(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/29038 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US01/31377

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/237,862 4 October 2000 (04.10.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/237,862 (CIP)
Filed on 4 October 2000 (04.10.2000)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HERRMANN, John, L.** [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **SHIMKETS, Richard, A.** [US/US]; 191 Leete Street, West Haven, CT 06516 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Finantial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/29038 A2

(54) Title: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME AND ANTIBODIES DIRECTED AGAINST THESE PROTEINS

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME AND ANTIBODIES DIRECTED AGAINST THESE PROTEINS

FIELD OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded thereby, and antibodies directed against the polypeptides.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, NOV4, NOV5, and NOV6 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, and 11. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, and 12. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, and 11.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, and 11) or a complement of said oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, and 12). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic

acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, Cancer, Leukodystrophies, Breast cancer, Ovarian cancer, Prostate cancer, Uterine cancer, Hodgkin disease, Adenocarcinoma, Adrenoleukodystrophy,Cystitis, incontinence, Von Hippel-Lindau (VHL) syndrome, hypercalceimia, Endometriosis, Hirschsprung's disease, Crohn's Disease, Appendicitis, Cirrhosis, Liver failure, Wolfram Syndrome, Smith-Lemli-Opitz syndrome, Retinitis pigmentosa, Leigh syndrome; Congenital Adrenal Hyperplasia, Xerostomia; tooth decay and other dental problems; Inflammatory bowel disease, Diverticular disease, fertility, Infertility, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis , atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, obesity, Diabetes Insipidus and Mellitus with Optic Atrophy and Deafness, Pancreatitis, Metabolic Dysregulation, transplantation recovery, Autoimmune disease, Systemic lupus erythematosus, asthma, arthritis, psoriasis, Emphysema, Scleroderma, allergy, ARDS, Immunodeficiencies, Graft vesus host, Alzheimer's disease, Stroke, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Multiple sclerosis,Ataxia-telangiectasia, Behavioral disorders, Addiction, Anxiety, Pain, Neurodegeneration, Muscular dystrophy,Lesch-Nyhan syndrome,Myasthenia gravis, schizophrenia, and other dopamine-dysfunctional states, levodopa-induced dyskinesias, alcoholism, pileptic seizures and other neurological disorders, mental depression, Cerebellar ataxia, pure; Episodic ataxia, type 2; Hemiplegic migraine, Spinocerebellar ataxia-6, Tuberous sclerosis, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Renal tubular acidosis, IgA nephropathy, and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof. By

way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

The invention further includes a method of using antibodies that are specific for a NOVx polypeptide to treat a disease. The method includes treating a patient with an effective amount of the antibody to block the mechamisn of their pathology. Pathologies that are blocked by the use of NOVX antibodies include metastatic potential and invasion in kidney and gastric tumors; cell growth and cell survival in colon, breast, liver and gastric tumors; cell growth and cell survival in colon, breast, liver and gastric tumors; metastasis in breast and brain tumors; metastasis and chemotherapy resistance in colon, gastric, ovarian and lung tumors; and angiogenesis and tumor growth in liver cancer.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	GMAC034209_A	1	2	UNC5-like
2	CG-SC29263825	3	4	Fat 2 (FAT2) cadherin related tumor suppressor like
3	CG-SC 17661211	5	6	orphan GPCR-like
4	CG-SC28471525	7	8	Slit-like
5	AC133 antigen	9	10	AC133 antigen-like
6	NM_012445	11	12	Spondin 2-like

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

NOV1 is homologous to a UNC5-like family of proteins. NOV1 could be used to treat metastatic potential and invasion. Therapeutic targeting of NOV1 with a monoclonal antibody is anticipated to limit or block the extent of metastatic potential and invasion in kidney, gastric, and various other tumors.

NOV2 is homologous to the Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor-like family of proteins. Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor has homology to the b-catenin binding regions of classical cadherin cytoplasmic tails and also ends with a PDZ domain-binding motif. Protocadherin regulates branching morphogenesis in the kidneys and lungs. Therefore, NOV2 has a role in cell growth and cell survival. Therapeutic targeting of NOV2 with a monoclonal antibody is anticipated to limit or block the extent of cell growth and cell survival in colon, breast, liver, gastric, and various other tumors.

NOV3 is homologous to a family of Orphan GPCR-like proteins. Because of its high homology to GPCRs and its containing GPCR 7 transmembrane domains, NOV3 is thought to be involved with cell growth and cell survival. Therapeutic targeting of NOV3 with a monoclonal antibody is anticipated to limit or block the extent of cell growth and cell survival in colon, breast, liver, gastric, and various other tumors.

NOV4 is homologous to the Slit-like family of proteins. NOV4 blocks Natriuretic peptide receptor proteins, possibly a receptor with ATP binding and Kinase activity. NOV4 is thought to be involved with metastatic potential. Therapeutic targeting of NOV4 with a monoclonal antibody is anticipated to limit or block the extent of metastasis and invasion in breast, brain, and various other tumors.

NOV5 is homologous to the AC133 Antigen-like family of proteins. NOV5 is thought to be involved in metastatic potential and chemotherapy resistance. Therapeutic targeting of AC133 with a monoclonal antibody is anticipated to limit or block the extent of metastasis and chemotherapy resistance in colon, gastric, ovarian, lung, and various other tumors.

NOV6 is homologous to the Spondin 2-like family of proteins. It is thought that NOV6 is involved with liver cancer. Therapeutic targeting of NOV6 with a monoclonal antibody is

anticipated to limit or block the extent of angiogenesis and tumor growth in liver, and various other cancers.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell proliferation, hematopoiesis, wound healing and angiogenesis. Antibodies specific for NOVX can be used to treat certain pathologies as detailed above.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A disclosed NOV1 nucleic acid of 2881 nucleotides (also referred to as GMAC034209_A) encoding a novel UNC5-like protein is shown in Table 1A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 87-89 and ending with a TGA codon at nucleotides 2784-2786. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 1A. The start and stop codons are in bold letters.

Table 1A. NOV1 nucleotide sequence (SEQ ID NO:1).

AGCTGGGGCTCCGGGCTGAGGCGCTAAAGCCGCCCTCCCGCCCGCGGGGCCCCGCGCCCGGCCCGCCCGCCT
GCCCGCCCGCGGCCATGGCCGTCCGGCCCGGCCTGTGGCCAGCGCTCCTGGGCATAGTCCTCGCCGCTTGGC
TCCGCGGCTCGGGTGCCCAGCAGAGTGCCACCGTGGCCAACCCAGTGCCTGGTGCCAACCCGGACCTGCTTC
CCCACTTCCTGGTGGAGCCCGAGGATGTGTACATCGTCAAGAACAAGCCAGTGCTGCTTGTGTGCAAGGCCG
TGCCCGCCACGCAGATCTTCTTCAAGTGCAACGGGGAGTGGGTGCGCCAGGTGGACCACGTGATCGAGCGCA
GCACAGACGGGAGCAGTGGTGAGCCGACCATGGAGGTCCGCATTAATGTCTCAAGGCAGCAGGTCGAGAAGG
TGTTCGGGCTGGAGGAATACTGGTGCCAGTGCGTGGCATGGAGCTCCTCGGGCACCACCAAGAGTCAGAAGG
CCTACATCCGCATAGCCAGATTGCGCAAGAACTTCGAGCAGGAGCCGCTGGCCAAGGAGGTGTCCCTGGAGC
AGGGCATCGTGCTGCCCTGCCGTCCACCGGAGGGCATCCCTCCAGCCGAGGTGGAGTGGCTCCGGAACGAGG
ACCTGGTGGACCCGTCCCTGGACCCCAATGTATACATCACGCGGGAGCACAGCCTGGTGGTGCGACAGGCCC
GCCTTGCTGACACGGCCAACTACACCTGCGTGGCCAAGAACATCGTGGCACGTCGCCGCAGCGCCTCCGCTG
CTGTCATCGTCTACGTGAACGGTGGGTGGTCGACGTGGACCGAGTGGTCCGTCTGCAGCGCCAGCTGTGGGC
GCGGCTGGCAGAAACGGAGCCGGAGCTGCACCAACCCGGCGCCTCTCAACGGGGGCGCTTTCTGTGAGGGGC
AGAATGTCCATGACCGCACCGTCTCCTCTCTGCTTGTCTCTGTGGACGGCAGCTGGAGCCCGTGGAGCAAGT
GGTCGGCCTGTGGGCTGGACTGCACCCACTGGCGGAGCCGTGAGTGCTCTGACCCAGCACCCCGCAACGGAG
GGGAGGAGTGCCAGGGCACTGACCTGGACACCCGCAACTGTACCAGTGACCTCTGTGTACACAGTGCTTCTG
GCCCTGAGGACGTGGCCCTCTATGTGGGCCTCATCGCCGTGGCCGTCTGCCTGGTCCTGCTGCTGCTTGTCC
TCATCCTCGTTTATTGCCGGAAGAAGGAGGGGCTGGACTCAGATGTGGCTGACTCGTCCATTCTCACCTCAG
GCTTCCAGCCCGTCAGCATCAAGCCCAGCAAAGCAGACAACCCCCATCTGCTCACCATCCAGCCGGACCTCA
GCACCACCACCACCTACCAGGGCAGTCTCTGTCCCCGGCAGGATGGGCCCAGCCCCAAGTTCCAGCTCACCA
ATGGGCACCTGCTCAGCCCCCTGGGTGGCGGCCGCCACACACTGCACCACAGCTCTCCCACCTCTGAGGCCG
AGGAGTTCGTCTCCCGCCTCTCCACCCAGAACTACTTCCGCTCCCTGCCCCGAGGCACCAGCAACATGACCT
ATGGGACCTTCAACTTCCTCGGGGGCCGGCTGATGATCCCTAATACAGGTATCAGCCTCCTCATCCCCCCAG
ATGCCATACCCCGAGGGAAGATCTATGAGATCTACCTCACGCTGCACAAGCCGGAAGACGTGAGGTTGCCCC
TAGCTGGCTGTCAGACCCTGCTGAGTCCCATCGTTAGCTGTGGACCCCCTGGCGTCCTGCTCACCCGGCCAG
TCATCCTGGCTATGGACCACTGTGGGGAGCCCAGCCCTGACAGCTGGAGCCTGCGCCTCAAAAAGCAGTCGT
GCGAGGGCAGCTGGGAGCAGGATGTGCTGCACCTGGGCGAGGAGGCGCCCTCCCACCTCTACTACTGCCAGC
TGGAGGCCAGTGCCTGCTACGTCTTCACCGAGCAGCTGGGCCGCTTTGCCCTGGTGGGAGAGGCCCTCAGCG

```
TGGCTGCCGCCAAGCGCCTCAAGCTGCTTCTGTTTGCGCCGGTGGCCTGCACCTCCCTCGAGTACAACATCC
GGGTCTACTGCCTGCATGACACCCACGATGCACTCAAGGAGGTGGTGCAGCTGGAGAAGCAGCTGGGGGGAC
AGCTGATCCAGGAGCCACGGGTCCTGCACTTCAAGGACAGTTACCACAACCTGCGCCTATCCATCCACGATG
TGCCCAGCTCCCTGTGGAAGAGTAAGCTCCTTGTCAGCTACCAGGAGATCCCCTTTTATCACATCTGGAATG
GCACGCAGCGGTACTTGCACTGCACCTTCACCCTGGAGCGTGTCAGCCCCAGCACTAGTGACCTGGCCTGCA
AGCTGTGGGTGTGGCAGGTGGAGGGCGACGGGCAGAGCTTCAGCATCAACTTCAACATCACCAAGGACACAA
GGTTTGCTGAGCTGCTGGCTCTGGAGAGTGAAGCGGGGGTCCCAGCCCTGGTGGGCCCCAGTGCCTTCAAGA
TCCCCTTCCTCATTCGGCAGAAGATAATTTCCAGCCTGGACCCACCCTGTAGGCGGGGTGCCGACTGGCGGA
CTCTGGCCCAGAAACTCCACCTGGACAGCCATCTCAGCTTCTTTGCCTCCAAGCCCAGCCCCACAGCCATGA
TCCTCAACCTGTGGGAGGCGCGGCACTTCCCCAACGGCAACCTCAGCCAGCTGGCTGCAGCAGTGGCTGGAC
TGGGCCAGCCAGACGCTGGCCTCTTCACAGTGTCGGAGGCTGAGTGCTGAGGCCGGCCAGGCCCGACACCTA
CACTCTCACCAGCTTTGGCACCCACCAAGGACAGGCAGAAGCCGGACAGGGGCCCTTCCCCACACCGGGGAG
A
```

In a search of public sequence databases, the NOV1 nucleic acid sequence, located on chromosome 13 has 1718 of 1725 bases (99%) identical to a *Homo sapiens* sequence similar to transmembrane receptor Unc5H1 from *Rattus norvegicus*. (gb:GENBANK-ID: gi|14781377|ref|XM_030300.1|). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the NOV1 BLAST analysis, *e.g.*, *Homo sapiens* sequence similar to transmembrane receptor Unc5H1 from *Rattus norvegicus*, matched the Query NOV1 sequence purely by chance is 0.0. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, e.g., http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-

complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. (Wootton and Federhen, Methods Enzymol 266:554-571, 1996).

The disclosed NOV1 polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 899 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1 is likely to be localized in the plasma membrane.

TaqMan data for NOV1 can be found below in Example 1. It indicates overexpression of NOV1 in kidney and gastric tumors.

Table 1B. Encoded NOV1 protein sequence (SEQ ID NO:2).

```
MAVRPGLWPALLGIVLAAWLRGSGAQQSATVANPVPGANPDLLPHFLVEPEDVYIVKNKPVLLVCKAVPATQ
IFFKCNGEWVRQVDHVIERSTDGSSGEPTMEVRINVSRQQVEKVFGLEEYWCQCVAWSSSGTTKSQKAYIRI
ARLRKNFEQEPLAKEVSLEQGIVLPCRPPEGIPPAEVEWLRNEDLVDPSLDPNVYITREHSLVVRQARLADT
ANYTCVAKNIVARRRSASAAVIVYVNGGWSTWTEWSVCSASCGRGWQKRSRSCTNPAPLNGGAFCEGQNVHD
RTVSSLLVSVDGSWSPWSKWSACGLDCTHWRSRECSDPAPRNGGEECQGTDLDTRNCTSDLCVHSASGPEDV
ALYVGLIAVAVCLVLLLLVLILVYCRKKEGLDSDVADSSILTSGFQPVSIKPSKADNPHLLTIQPDLSTTTT
YQGSLCPRQDGPSPKFQLTNGHLLSPLGGGRHTLHHSSPTSEAEEFVSRLSTQNYFRSLPRGTSNMTYGTFN
FLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTLLSPIVSCGPPGVLLTRPVILAM
DHCGEPSPDSWSLRLKKQSCEGSWEQDVLHLGEEAPSHLYYCQLEASACYVFTEQLGRFALVGEALSVAAAK
RLKLLLFAPVACTSLEYNIRVYCLHDTHDALKEVVQLEKQLGGQLIQEPRVLHFKDSYHNLRLSIHDVPSSL
WKSKLLVSYQEIPFYHIWNGTQRYLHCTFTLERVSPSTSDLACKLWVWQVEGDGQSFSINFNITKDTRFAEL
LALESEAGVPALVGPSAFKIPFLIRQKIISSLDPPCRRGADWRTLAQKLHLDSHLSFFASKPSPTAMILNLW
EARHFPNGNLSQLAAAVAGLGQPDAGLFTVSEAEC
```

A search of sequence databases reveals that the NOV1 amino acid sequence has 812 of 900 amino acid residues (90%) identical to, and 828 of 900 amino acid residues (91%) similar to the 898 amino acid residue transmembrane receptor Unc5H1 [Rattus norvegicus] (GenBank Acc. No.: gi|11559980|ref|NP_071542.1|) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV1 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 1C.

Table 1C. BLAST results for NOV1

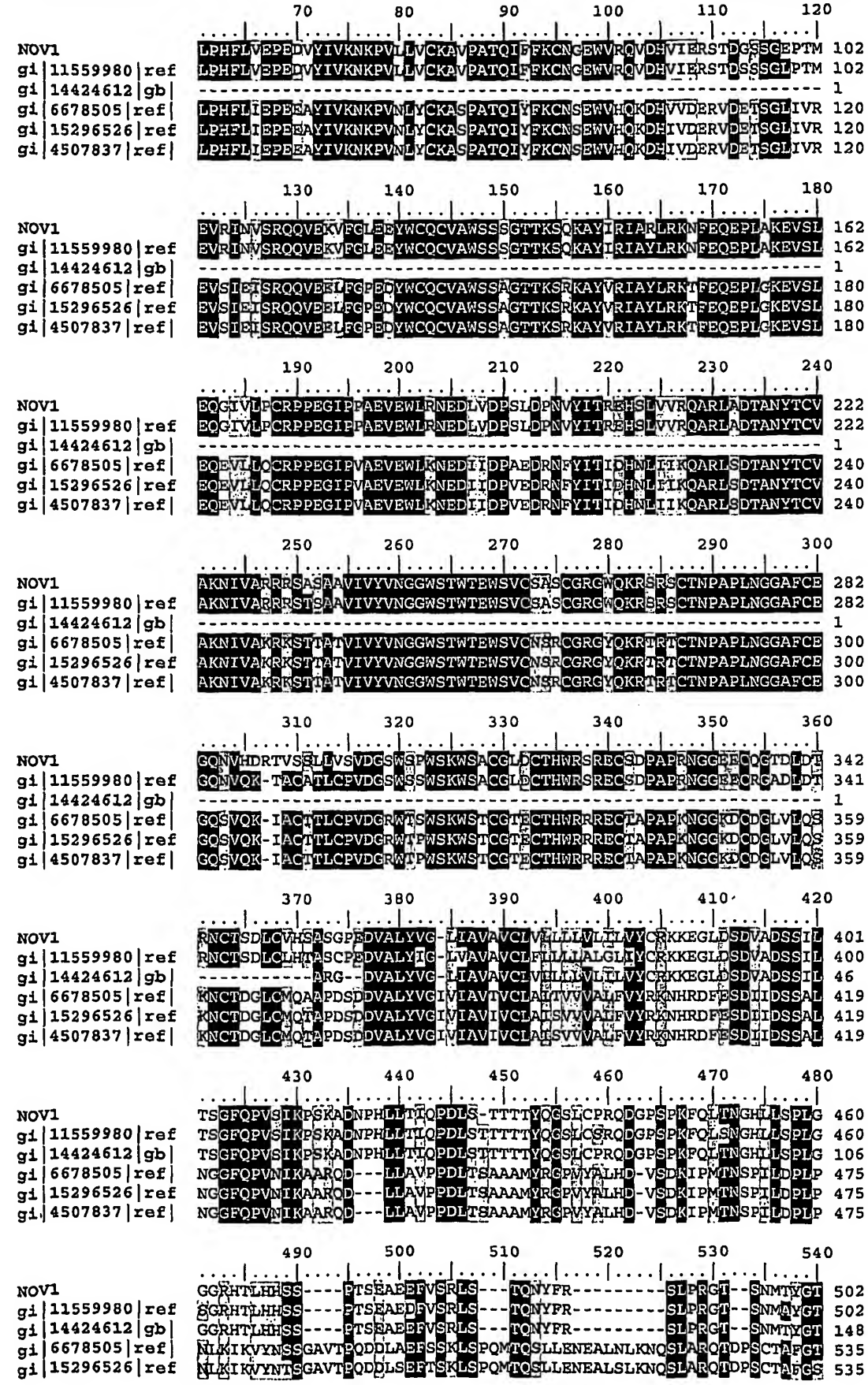
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11559980|ref|NP_071542.1|	transmembrane receptor Unc5H1 [Rattus norvegicus]	898	812/900 (90%)	828/900 (91%)	0.0
gi|14424612|gb|AAH09333.1|AAH09333 (BC009333)	Similar to transmembrane receptor Unc5H1 [Homo sapiens]	544	506/542 (93%)	506/542 (93%)	0.0
gi|6678505|ref|NP_033498.1|	UNC-5 homolog (C. elegans) 3 [Mus musculus]	931	490/913 (53%)	631/913 (68%)	e-161
gi|15296526|ref|XP_042940.2|	unc5 (C.elegans homolog) c [Homo sapiens]	931	483/913 (52%)	625/913 (67%)	e-160
gi|4507837|ref|NP_003719.1|	unc5 (C.elegans homolog) c; homolog of C. elegans transmembrane receptor Unc5 [Homo sapiens]	931	482/913 (52%)	624/913 (67%)	e-158

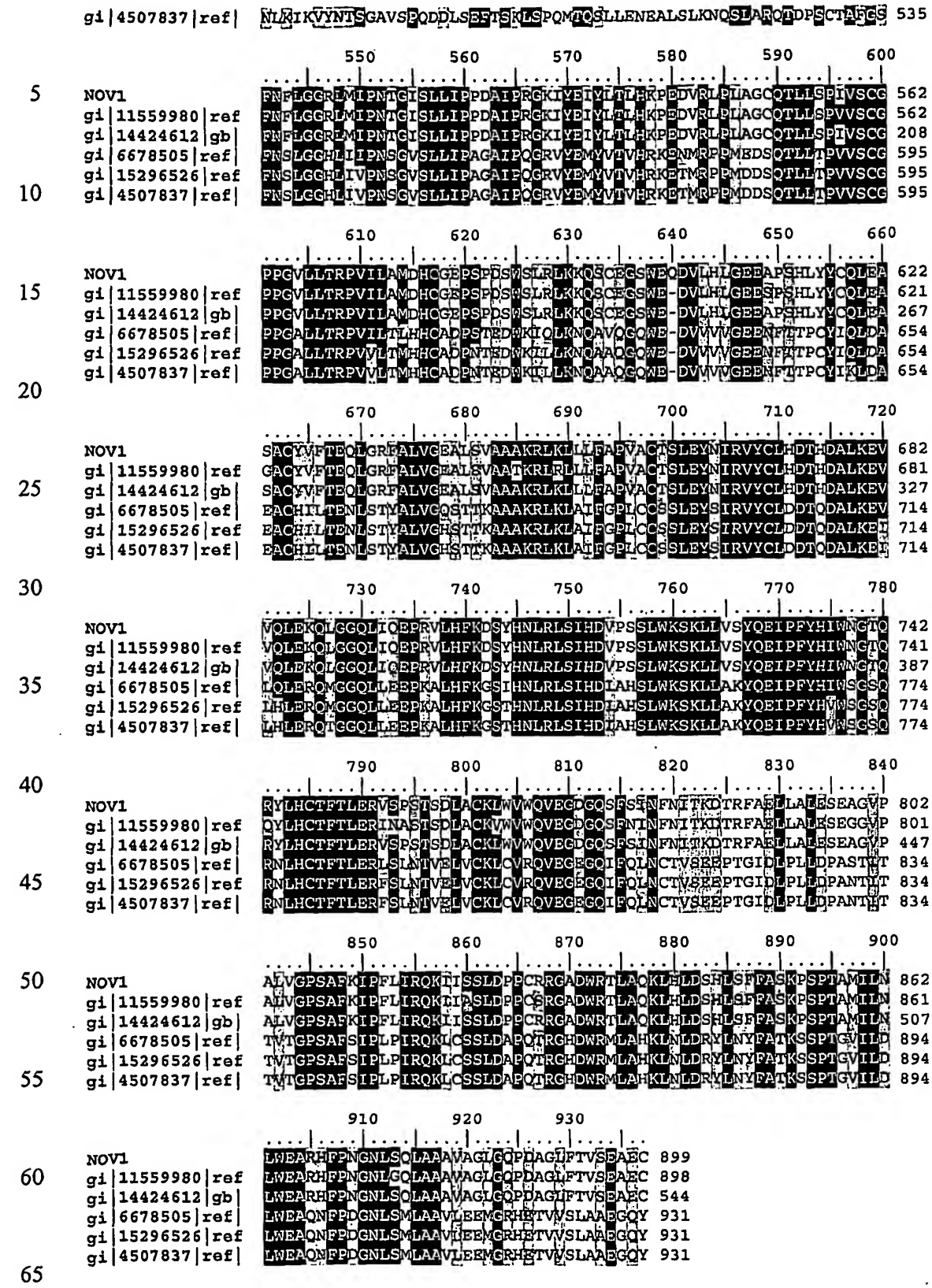
The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 1D. In the ClustalW alignment of the NOV1 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 1D. ClustalW Analysis of NOV1

1) Novel NOV1 (SEQ ID NO:2)
2) gi|11559980|ref|NP_071542.1| transmembrane receptor Unc5H1 [Rattus norvegicus] (SEQ ID NO:13)
3) gi|14424612|gb|AAH09333.1|AAH09333 (BC009333) Similar to transmembrane receptor Unc5H1 [Homo sapiens] (SEQ ID NO:14)
4) gi|6678505|ref|NP_033498.1| UNC-5 homolog (C. elegans) 3 [Mus musculus] (SEQ ID NO:15)
5) gi|15296526|ref|XP_042940.2| unc5 (C.elegans homolog) c [Homo sapiens] (SEQ ID NO:16)
6) gi|4507837|ref|NP_003719.1| unc5 (C.elegans homolog) c; homolog of C. elegans transmembrane receptor Unc5 [Homo sapiens] (SEQ ID NO:17)

```
                                10        20        30        40        50        60
                       ....|....|....|....|....|....|....|....|....|....|....|....|
NOV1                   -------MAVRPGLWP----ALLGIVLAAWLRGSGAQQSAT-------VANPVPGANPDL 42
gi|11559980|ref        -------MAVRPGLWP----VLLGIVLAAWLRGSGAQQSAT-------VANPVPGANPDL 42
gi|14424612|gb|        ------------------------------------------------------------ 1
gi|6678505|ref|       MRKGLRATAARCGLGLGYLLQMLVLPALALLSASGTGSAAQQDDEFFHELPETPPSDPPEP 60
gi|15296526|ref       MRKGLRATAARCGLGLGYLLQMLVLPALALLSASGTGSAAQQDDDFFHELPETPPSDPPEP 60
gi|4507837|ref|       MRKGLRATAARCGLGLGYLLQMLVLPALALLSASGTGSAAQQDDDFFHELPETPPSDPPEP 60
```


70 80 90 100 110 120
NOV1 LPHFLVEPEDVYIVKNKPVLLVCKAVPATQIFFKCNGEWVRQVDHVIERSTDGSSGEPTM 102
gi|11559980|ref LPHFLVEPEDVYIVKNKPVLLVCKAVPATQIFFKCNGEWVRQVDHVIERSTDSSSGLPTM 102
gi|14424612|gb| ---------- 1
gi|6678505|ref| LPHFLIEPEEAYIVKNKPVNLYCKASPATQIYFKCNSEWVHQKDHVVDERVDETSGLIVR 120
gi|15296526|ref LPHFLIEPEEAYIVKNKPVNLYCKASPATQIYFKCNSEWVHQKDHIVDERVDETSGLIVR 120
gi|4507837|ref| LPHFLIEPEEAYIVKNKPVNLYCKASPATQIYFKCNSEWVHQKDHIVDERVDETSGLIVR 120
130 140 150 160 170 180
NOV1 EVRINVSRQQVEKVFGLEEYWCQCVAWSSSGTTKSQKAYIRIARLRKNFEQEPLAKEVSL 162
gi|11559980|ref EVRINVSRQQVEKVFGLEEYWCQCVAWSSSGTTKSQKAYIRIAYLRKNFEQEPLAKEVSL 162
gi|14424612|gb| ---------- 1
gi|6678505|ref| EVSIEISRQQVEELFGPEDYWCQCVAWSSAGTTKSRKAYVRIAYLRKTFEQEPLGKEVSL 180
gi|15296526|ref EVSIEISRQQVEELFGPEDYWCQCVAWSSAGTTKSRKAYVRIAYLRKTFEQEPLGKEVSL 180
gi|4507837|ref| EVSIEISRQQVEELFGPEDYWCQCVAWSSAGTTKSRKAYVRIAYLRKTFEQEPLGKEVSL 180
190 200 210 220 230 240
NOV1 EQGIVLPCRPPEGIPPAEVEWLRNEDLVDPSLDPNVYITREHSLVVRQARLADTANYTCV 222
gi|11559980|ref EQGIVLPCRPPEGIPPAEVEWLRNEDLVDPSLDPNVYITREHSLVVRQARLADTANYTCV 222
gi|14424612|gb| ---------- 1
gi|6678505|ref| EQEVLLQCRPPEGIPVAEVEWLKNEDIIDPAEDRNFYITIDHNLIIKQARLSDTANYTCV 240
gi|15296526|ref EQEVLLQCRPPEGIPVAEVEWLKNEDIIDPVEDRNFYITIDHNLIIKQARLSDTANYTCV 240
gi|4507837|ref| EQEVLLQCRPPEGIPVAEVEWLKNEDIIDPVEDRNFYITIDHNLIIKQARLSDTANYTCV 240
250 260 270 280 290 300
NOV1 AKNIVARRRSASAAVIVYVNGGWSTWTEWSVCSASCGRGWQKRSRSCTNPAPLNGGAFCE 282
gi|11559980|ref AKNIVARRRSTSAAVIVYVNGGWSTWTEWSVCSASCGRGWQKRSRSCTNPAPLNGGAFCE 282
gi|14424612|gb| ---------- 1
gi|6678505|ref| AKNIVAKRKSTIATVIVYVNGGWSTWTEWSVCNSRCGRGYQKRTRTCTNPAPLNGGAFCE 300
gi|15296526|ref AKNIVAKRKSTIATVIVYVNGGWSTWTEWSVCNSRCGRGYQKRTRTCTNPAPLNGGAFCE 300
gi|4507837|ref| AKNIVAKRKSTIATVIVYVNGGWSTWTEWSVCNSRCGRGYQKRTRTCTNPAPLNGGAFCE 300
310 320 330 340 350 360
NOV1 GQNVHDRTVSSLLVSVDGSWSPWSKWSACGLDCTHWRSRECSDPAPRNGGEECQGTDLDT 342
gi|11559980|ref GQNVQK-TACATLCPVDGSWSSWSKWSACGLDCTHWRSRECSDPAPRNGGEECRGADLDT 341
gi|14424612|gb| ---------- 1
gi|6678505|ref| GQSVQK-IACTTLCPVDGRWTSWSKWSICGTECTHWRRRECTAPAPKNGGKDCDGLVLQS 359
gi|15296526|ref GQSVQK-IACTTLCPVDGRWTPWSKWSICGTECTHWRRRECTAPAPKNGGKDCDGLVLQS 359
gi|4507837|ref| GQSVQK-IACTTLCPVDGRWTPWSKWSICGTECTHWRRRECTAPAPKNGGKDCDGLVLQS 359
370 380 390 400 410 420
NOV1 RNCTSDLCVHSASGPEDVALYVG-LIAVAVCLVLLLVLILVYCRKKEGLDSDVADSSIL 401
gi|11559980|ref RNCTSDLCLHHTASCPEDVALYIG-LVAVAVCLFLLLLALGLIYCRKKEGLDSDVADSSIL 400
gi|14424612|gb| ----------ARG--DVALYVG-LIAVAVCLVLLLLVLILVYCRKKEGLDSDVADSSIL 46
gi|6678505|ref| KNCTDGLCMQAAPDSDDVALYVGIVIAVIVCLAITVVVALFVYRKNHRDFFSDIIDSSAL 419
gi|15296526|ref KNCTDGLCMQTAPDSDDVALYVGIVIAVIVCLAISVVVALFVYRKNHRDFFSDIIDSSAL 419
gi|4507837|ref| KNCTDGLCMQTAPDSDDVALYVGIVIAVIVCLAISVVVALFVYRKNHRDFFSDIIDSSAL 419
430 440 450 460 470 480
NOV1 TSGFQPVSIKPSKADNPHLLTIQPDLS-TTTTYQGSLCPRQDGPSPKFQLINGHLLSPLG 460
gi|11559980|ref TSGFQPVSIKPSKADNPHLLTIQPDLSTTTTTYQGSLCSRQDGPSPKFQLSNGHLLSPLG 460
gi|14424612|gb| TSGFQPVSIKPSKADNPHLLTIQPDLSTTTTTYQGSLCPRQDGPSPKFQLTNGHLLSPLG 106
gi|6678505|ref| NGGFQPVNIKAARQD---LLAVPPDLTSAAAMYRGPVYALHD-VSDKIPMTNSPILDPLP 475
gi|15296526|ref NGGFQPVNIKAARQD---LLAVPPDLTSAAAMYRGPVYALHD-VSDKIPMTNSPILDPLP 475
gi|4507837|ref| NGGFQPVNIKAARQD---LLAVPPDLTSAAAMYRGPVYALHD-VSDKIPMTNSPILDPLP 475
490 500 510 520 530 540
NOV1 GGRHTLHHSS----PTSEAEEFVSRLS---TQNYFR---------SLPRGT--SNMTYGT 502
gi|11559980|ref GGRHTLHHSS----PTSEAEDFVSRLS---TQNYFR---------SLPRGT--SNMAYGT 502
gi|14424612|gb| GGRHTLHHSS----PTSEAEEFVSRLS---TQNYFR---------SLPRGT--SNMTYGT 148
gi|6678505|ref| NLKIKVYNSSGAVTPQDDLAEFSSKLSPQMTQSLLENEALNLKNQSLARQTDPSCTAFGT 535
gi|15296526|ref NLKIKVYNTSGAVTPQDDLSEFTSKLSPQMTQSLLENEALSLKNQSLARQTDPSCTAFGS 535


gi|4507837|ref| NLKIKVYNTSGAVSPQDDLSETSKLSPQMTQSLLENEALSLKNQSLARQTDPSCTAFCS 535
550 560 570 580 590 600
NOV1 FNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTLLSPVVSCG 562
gi|11559980|ref FNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTLLSPVVSCG 562
gi|14424612|gb| FNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTLLSPIVSCG 208
gi|6678505|ref| FNSLGGHLIIPNSGVSLLIPAGAIPQGRVYEMYVTVHRKSNMRPPMEDSQTLLTPVVSCG 595
gi|15296526|ref FNSLGGHLIVPNSGVSLLIPAGAIPQGRVYEMYVTVHRKETMRPPMDDSQTLLTPVVSCG 595
gi|4507837|ref| FNSLGGHLIVPNSGVSLLIPAGAIPQGRVYEMYVTVHRKETMRPPMDDSQTLLTPVVSCG 595
610 620 630 640 650 660
NOV1 PPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEGSWEQDVLHLGEEAPSHLYYCQLEA 622
gi|11559980|ref PPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEGSWE-DVLHLGEESPSHLYYCQLEA 621
gi|14424612|gb| PPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEGSWE-DVLHLGEEAPSHLYYCQLEA 267
gi|6678505|ref| PPGALLTRPVILTLHHCADPSTEDWKIQLKNQAVQGQWE-DVVVVGEENFTTPCYIQLDA 654
gi|15296526|ref PPGALLTRPVVLTMHHCADPNTEDWKLLLKNQAAQGQWE-DVVVVGEENFTTPCYIQLDA 654
gi|4507837|ref| PPGALLTRPVVLTMHHCADPNTEDWKLLLKNQAAQGQWE-DVVVVGEENFTTPCYIKLDA 654
670 680 690 700 710 720
NOV1 SACYVFTEQLGRFALVGEALSVAAAKRLKLLLFAPVACTSLEYNIRVYCLHDTHDALKEV 682
gi|11559980|ref GACYVFTEQLGRFALVGEALSVAAAKRLRLLLFAPVACTSLEYNIRVYCLHDTHDALKEV 681
gi|14424612|gb| SACYVFTEQLGRFALVGEALSVAAAKRLKLLLFAPVACTSLEYNIRVYCLHDTHDALKEV 327
gi|6678505|ref| EACHILTENLSTYALVGQSTTKAAAKRLKLAIFGPLCCSSLEYSIRVYCLDDTQDALKEV 714
gi|15296526|ref EACHILTENLSTYALVGHSTTKAAAKRLKLAIFGPLCCSSLEYSIRVYCLDDTQDALKEI 714
gi|4507837|ref| EACHILTENLSTYALVGHSTTKAAAKRLKLAIFGPLCCSSLEYSIRVYCLDDTQDALKEI 714
730 740 750 760 770 780
NOV1 VQLEKQLGGQLIQEPRVLHFKDSYHNLRLSIHDVPSSLWKSKLLVSYQEIPFYHIWNGTQ 742
gi|11559980|ref VQLEKQLGGQLIQEPRVLHFKDSYHNLRLSIHDVPSSLWKSKLLVSYQEIPFYHIWNGTQ 741
gi|14424612|gb| VQLEKQLGGQLIQEPRVLHFKDSYHNLRLSIHDVPSSLWKSKLLVSYQEIPFYHIWNGTQ 387
gi|6678505|ref| LQLERQMGGQLLEEPKALHFKGSIHNLRLSIHDLAHSLWKSKLLAKYQEIPFYHIWSGSQ 774
gi|15296526|ref IHLERQMGGQLLEEPKALHFKGSTHNLRLSIHDLAHSLWKSKLLAKYQEIPFYHVWSGSQ 774
gi|4507837|ref| LHLERQTGGQLLEEPKALHFKGSTHNLRLSIHDLAHSLWKSKLLAKYQEIPFYHVWSGSQ 774
790 800 810 820 830 840
NOV1 RYLHCTFTLERVSPSTSDLACKLWVWQVEGDGQSFSINFNITKDTRFAELLALESEAGVP 802
gi|11559980|ref QYLHCTFTLERINASTSDLACKVWVWQVEGDGQSFNINFNITKDTRFAELLALESEGGVP 801
gi|14424612|gb| RYLHCTFTLERVSPSTSDLACKLWVWQVEGDGQSFSINFNITKDTRFAELLALESEAGVP 447
gi|6678505|ref| RNLHCTFTLERLSLNTVELVCKLCVRQVEGEGQIFQLNCTVSEEPTGIDLPLLDPASTTT 834
gi|15296526|ref RNLHCTFTLERFSLNTVELVCKLCVRQVEGEGQIFQLNCTVSEEPTGIDLPLLDPANTTT 834
gi|4507837|ref| RNLHCTFTLERFSLNTVELVCKLCVRQVEGEGQIFQLNCTVSEEPTGIDLPLLDPANTTT 834
850 860 870 880 890 900
NOV1 ALVGPSAFKIPFLIRQKLISSLDPPCRRGADWRTLAQKLELDSHLSFFASKPSPTAMILN 862
gi|11559980|ref ALVGPSAFKIPFLIRQKLIASLDPPCSRGADWRTLAQKLHLDSHLSFFASKPSPTAMILN 861
gi|14424612|gb| ALVGPSAFKIPFLIRQKLISSLDPPCRRGADWRTLAQKLHLDSHLSFFASKPSPTAMILN 507
gi|6678505|ref| TVTGPSAFSIPLPIRQKLCSSLDAPQTRGHDWRMLAHKLNLDRYLNYFATKSSPTGVILD 894
gi|15296526|ref TVTGPSAFSIPLPIRQKLCSSLDAPQTRGHDWRMLAHKLNLDRYLNYFATKSSPTGVILD 894
gi|4507837|ref| TVTGPSAFSIPLPIRQKLCSSLDAPQTRGHDWRMLAHKLNLDRYLNYFATKSSPTGVILD 894
910 920 930
NOV1 LWEARHFPNGNLSQLAAAVAGLGQPDAGLFTVSEAEC 899
gi|11559980|ref LWEARHFPNGNLGQLAAAVAGLGQPDAGLFTVSEAEC 898
gi|14424612|gb| LWEARHFPNGNLSQLAAAVAGLGQPDAGLFTVSEAEC 544
gi|6678505|ref| LWEAQNFPDGNLSMLAAVLEEMGRHETVVSLAAEGQY 931
gi|15296526|ref LWEAQNFPDGNLSMLAAVLEEMGRHETVVSLAAEGQY 931
gi|4507837|ref| LWEAQNFPDGNLSMLAAVLEEMGRHETVVSLAAEGQY 931

The presence of identifiable domains in NOV1, as well as all other NOVX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/ interpro). DOMAIN results for NOV1 as disclosed in Tables 1E-IL, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 1E and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (|) and "strong" semi-conserved residues are indicated by grey shading or by the sign (+). The "strong" group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Tables 1E-1L list the domain description from DOMAIN analysis results against NOV1. This indicates that the NOV1 sequence has properties similar to those of other proteins known to contain this domain.

Table 1E. Domain Analysis of NOV1

gnl|Smart|smart00218, ZU5, Domain present in ZO-1 and Unc5-like netrin receptors; Domain of unknown function. (SEQ ID NO:42)
CD-Length = 51 residues, 100.0% aligned
Score = 49.7 bits (117), Expect = 7e-07

```
Query:  495  TSNMTYGTFNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTL  554
             | +  |||+  ||||  |  |+ |+||| |||+|  |  || +|       ||   +||
Sbjct:  1    PSFLVSGTFDARGGRLRGPRTGVRLIIPPGAIPQGTRYTCYLVVHDKLSTPPPLEEGETL  60

Query:  555  LSPIVSCGPPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEG  598
             |||+| ||| | |  ||||| + ||    |  | + | +    |
Sbjct:  61   LSPVVECGPHGALFLRPVILEVPHCASLRPRDWEIVLLRSENGG  104
```

Table 1F. Domain Analysis of NOV1

gnl|Smart|smart00082, LRRCT, Leucine rich repeat C-terminal domain. (SEQ ID NO:43)
CD-Length = 104 residues, 100.0% aligned
Score = 152 bits (383), Expect = 1e-37

```
Query:  495  TSNMTYGTFNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTL  554
             | +  |||+  ||||  |  |+ |+||| |||+|  |  || +|       ||   +||
Sbjct:  1    PSFLVSGTFDARGGRLRGPRTGVRLIIPPGAIPQGTRYTCYLVVHDKLSTPPPLEEGETL  60

Query:  555  LSPIVSCGPPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEG  598
             |||+| ||| | |  ||||| + ||    |  | + | +    |
Sbjct:  61   LSPVVECGPHGALFLRPVILEVPHCASLRPRDWEIVLLRSENGG  104
```

Table 1G. Domain Analysis of NOV1

gnl|Pfam|pfam00791, ZU5, ZU5 domain. Domain present in ZO-1 and Unc5-like netrin receptors Domain of unknown function. (SEQ ID NO:44)
CD-Length = 104 residues, 100.0% aligned
Score = 150 bits (378), Expect = 4e-37

```
Query:  495  TSNMTYGTFNFLGGRLMIPNTGISLLIPPDAIPRGKIYEIYLTLHKPEDVRLPLAGCQTL  554
Sbjct:  1    SGFLVSGTFDARGGRLRGPRTGVRLIIPPGAIPQGTRYTCYLVVHDKLSTPPPLEEGETL  60

Query:  555  LSPIVSCGPPGVLLTRPVILAMDHCGEPSPDSWSLRLKKQSCEG  598
Sbjct:  61   LSPVVECGPHGALFLRPVILEVPHCASLRPRDWELVLLRSENGG  104
```

Table 1H. Domain Analysis of NOV1

gnl|Smart|smart00005, DEATH, DEATH domain, found in proteins involved in cell death (apoptosis).; Alpha-helical domain present in a variety of proteins with apoptotic functions. Some (but not all) of these domains form homotypic and heterotypic dimers.. (SEQ ID NO:45)
CD-Length = 96 residues, 91.7% aligned
Score = 61.6 bits (148), Expect = 2e-10

```
Query:  806  GPSAFKIPFLIRQKIISSLDPPCRRGADWRTLAQKLHL-DSHLSFFASKPS-----PTAM  859
Sbjct:  1    PPGAASLTELTREKLAKLLDHD--LGDDWRELARKLGLSEADIDQIETESPRDLAEQSYQ  58

Query:  860  ILNLWEARHFPNGNLSQLAAAVAGLGQPDA  889
Sbjct:  59   LLRLWEQREGKNATLGTLLEALRKMGRDDA  88
```

Table 1I. Domain Analysis of NOV1

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta. (SEQ ID NO:46)
CD-Length = 51 residues, 84.3% aligned
Score = 56.2 bits (134), Expect = 8e-09

```
Query:  245  WSTWTEWSVCSASCGRGWQKRSRSCTNPAPLNGGAFCEGQNVHDR  289
Sbjct:  1    WGEWSEWSPCSVTCGGGVQTRTRCCNPPP--NGGGPCTGPDTETR  43
```

Table 1J. Domain Analysis of NOV1

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta. (SEQ ID NO:46)
CD-Length = 51 residues, 98.0% aligned
Score = 49.7 bits (117), Expect = 7e-07

```
Query:  302  WSPWSKWSACGLDCTH-WRSRECSDPAPRNGGEECQGTDLDTRNCTSDLC  350
Sbjct:  1    WGEWSEWSPCSVTCGGGVQTRTRCCNPPPNGGGPCTGPDTETRACNEQPC  50
```

Table 1K. Domain Analysis of NOV1

gnl|Pfam|pfam00531, death, Death domain (SEQ ID NO:47)
CD-Length = 83 residues, 90.4% aligned
Score = 52.8 bits (125), Expect = 9e-08

```
Query:  818  QKIISSLDPPCRRGADWRTLAQKLHL-DSHLSFFASKP----SPTAMILNLWEARHFPNG  872
             +    || |   | ||| ||+|| | +  +     +      |||  +|+||| |   |
Sbjct:  1    RELCKLLDDP--LGRDWRRLARKLGLSEEEIDQIEHENPRLASPTYQLLDLWEQRGGKNA  58

Query:  873  NLSQLAAAVAGLGQPDA  889
                 |  |+  +|+ ||
Sbjct:  59   TVGTLLEALRKMGRDDA  75
```

Table 1L. Domain Analysis of NOV1

gnl|Smart|smart00409, IG, Immunoglobulin (SEQ ID NO:48)
CD-Length = 86 residues, 79.1% aligned
Score = 44.3 bits (103), Expect = 3e-05

```
Query:  159  EVSLEQGIVLPCRPPEGIPPAEVEWLRNEDLVDPSLDPNVYITRE---HSLVVRQARLAD  215
             |  +  +|  |     | ||  | | +       +       +|     +| +      |
Sbjct:  5    TVKEGESVTLSCEAS-GNPPPTVTWYKQ-GGKLLAESGRFSVSRSGGNSTLTISNVTPED  62

Query:  216  TANYTCVAKN  225
             +  ||| | |
Sbjct:  63   SGTYTCAATN  72
```

Murine netrin-3 protein binds to netrin receptors of the DCC (deleted in colorectal cancer) family [DCC and neogenin] and the UNC5 family (UNC5H1, UNC5H2 and UNC5H3). *C elegans* Unc5 and murine unc5hr homolog are involved in cell migration during cerebellum development, inducing repulsion in axon guidance through its cytoplasmic tail, and are expressed in brain, fetal heart.

The disclosed NOV1 nucleic acid of the invention encoding a UNC5 -like protein includes the nucleic acid whose sequence is provided in Table 1A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 1A while still encoding a protein that maintains its UNC5-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least

in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 30% percent of the bases may be so changed.

The disclosed NOV1 protein of the invention includes the UNC5-like protein whose sequence is provided in Table 1B or 1E. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 1B or 1E while still encoding a protein that maintains its UNC5 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 48% percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this UNC5-like protein (NOV1) may function as a member of a "UNC5 family". Therefore, the NOV1 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here. NOV1 could be used to treat metastatic potential and invasion. Therapeutic targeting of NOV1 with a monoclonal antibody is anticipated to limit or block the extent of metastatic potential and invasion in kidney and gastric tumors.

NOV1 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV1 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders . These antibodies can also be used to treat certain pathological conditions as detailed above.

NOV2

A disclosed NOV2 nucleic acid of 14536 nucleotides (also referred to as CG-SC29263825 GenBank #AF231022) encoding a novel protocadherin Fat 2 (FAT2) cadherin related tumor suppressor like protein is shown in Table 2A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 14-16 and ending with a TAG codon at nucleotides 13061-13063. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 2A, and the start and stop codons are in bold letters.

Table 2A. NOV2 nucleotide sequence (SEQ ID NO:3).

GGAGTTTTCCACC**ATG**ACTATTGCCCTGCTGGGTTTTGCCATATTCTTGCTCCATTGTGCGACCTGTGAGAA
GCCTCTAGAAGGGATTCTCTCCTCCTCTGCTTGGCACTTCACACACTCCCATTACAATGCCACCATCTATGA
AAATTCTTCTCCCAAGACCTATGTGGAGAGCTTCGAGAAAATGGGCATCTACCTCGCGGAGCCACAGTGGGC
AGTGAGGTACCGGATCATCTCTGGGGATGTGGCCAATGTATTTAAAACTGAGGAGTATGTGGTGGGCAACTT
CTGCTTCCTAAGAATAAGGACAAAGAGCAGCAACACAGCTCTTCTGAACAGAGAGGTGCGAGACAGCTACAC
CCTCATCATCCAAGCCACAGAGAAGACCTTGGAGTTGGAAGCTTTGACCCGTGTGGTGGTCCACATCCTGGA
CCAGAATGACCTGAAGCCTCTCTTCTCTCCACCTTCGTACAGAGTCACCATCTCTGAGGACATGCCCCTGAA
GAGCCCCATCTGCAAGGTGACTGCCACAGATGCTGATCTAGGCCAGAATGCTGAGTTCTATTATGCCTTTAA
CACAAGGTCAGAGATGTTTGCCATCCATCCCACCAGCGGTGTGGTCACTGTGGCTGGGAAGCTTAACGTCAC
CTGGCGAGGAAAGCATGAGCTCCAGGTGCTAGCTGTGGACCGCATGCGGAAAATCTCTGAGGGCAATGGGTT
TGGCAGCCTGGCTGCACTTGTGGTTCATGTGGAGCCTGCCCTCAGGAAGCCCCCAGCCATTGCTTCGGTGGT
GGTGACTCCACCAGACAGCAATGATGGTACCACCTATGCCACTGTACTGGTCGATGCAAATAGCTCAGGAGC
TGAAGTGGAGTCAGTGGAAGTTGTTGGTGGTGACCCTGGAAAGCACTTCAAAGCCATCAAGTCTTATGCCCG
GAGCAATGAGTTCAGTTTGGTGTCTGTCAAAGACATCAACTGGATGGAGTACCTTCATGGGTTCAACCTCAG
CCTCCAGGCCAGGAGTGGGAGCGGCCCTTATTTTTATTCCCAGATCAGGGGCTTTCACCTACCACCTTCCAA
ACTGTCTTCCCTCAAATTCGAGAAGGCTGTTTACAGAGTGCAGCTTAGTGAGTTTTCCCCTCCTGGCAGCCG
CGTGGTGATGGTGAGAGTCACCCCAGCCTTCCCCAACCTGCAGTATGTTCTAAAGCCATCTTCAGAGAATGT
AGGATTTAAACTTAATGCTCGAACTGGGTTGATCACCACCACAAAGCTCATGGACTTCCACGACAGAGCCCA
CTATCAGCTACACATCAGAACCTCACCGGGCCAGGCCTCCACCGTGGTGGTCATTGACATTGTGGACTGCAA
CAACCATGCCCCCCTCTTCAACAGGTCTTCCTATGATGGTACCTTGGATGAGAACATCCCTCCAGGCACCAG
TGTTTTGGCTGTGACTGCCACTGACCGGGATCATGGGGAAAATGGATATGTCACCTATTCCATTGCTGGACC
AAAAGCTTTGCCATTTTCTATTGACCCCTACCTGGGGATCATCTCCACCTCCAAACCCATGGACTATGAACT
CATGAAAAGAATTTATACCTTCCGGGTAAGAGCATCAGACTGGGGATCCCCTTTTCGCCGGGAGAAGGAAGT
GTCCATTTTTCTTCAGCTCAGGAACTTGAATGACAACCAGCCTATGTTTGAAGAAGTCAACTGTACAGGGTC
TATCCGCCAAGACTGGCCAGTAGGGAAATCGATAATGACTATGTCAGCCATAGATGTGGATGAGCTTCAGAA
CCTAAAATACGAGATTGTATCAGGCAATGAACTAGAGTATTTTGATCTAAATCATTTCTCCGGAGTGATATC
CCTCAAACGCCCTTTTATCAATCTTACTGCTGGTCAACCCACCAGTTATTCCCTGAAGATTACAGCCTCAGA
TGGCAAAAACTATGCCTCACCCACAACTTTGAATATTACTGTGGTGAAGGACCCTCATTTTGAAGTTCCTGT
AACATGTGATAAAACAGGGGTATTGACACAATTCACAAAGACTATCCTCCACTTTATTGGGCTTCAGAACCA
GGAGTCCAGTGATGAGGAATTCACTTCTTTAAGCACATATCAGATTAATCATTACACCCCACAGTTTGAGGA
CCACTTCCCCCAATCCATTGATGTCCTTGAGAGTGTCCCTATCAACACCCCCTTGGCCCGCCTAGCAGCCAC
TGACCCTGATGCTGGTTTTAATGGCAAACTGGTCTATGTGATTGCAGATGGCAATGAGGAGGGCTGCTTTGA
CATAGAGCTGGAGACAGGGCTGCTCACTGTAGCTGCTCCCTTGGACTATGAAGCCACCAATTTCTACATCCT
CAATGTAACAGTATATGACCTGGGCACACCCCAGAAGTCCTCCTGGAAGCTGCTGACAGTGAATGTGAAAGA
CTGGAATGACAACGCACCCAGATTTCCTCCCGGTGGGTACCAGTTAACCATCTCGGAGGACACAGAAGTTGG
AACCACAATTGCAGAGCTGACAACCAAAGATGCTGACTCGGAAGACAATGGCAGGGTTCGCTACACCCTGCT
AAGTCCCACAGAGAAGTTCTCCCTCCACCCTCTCACTGGGGAACTGGTTGTTACAGGACACCTGGACCGCGA
ATCAGAGCCTCGGTACATACTCAAGGTGGAGGCCAGGGATCAGCCCAGCAAAGGCCACCAGCTCTTCTCTGT
CACTGACCTGATAATCACATTGGAGGATGTCAACGACAACTCTCCCCAGTGCATCACAGAACACAACAGGCT
GAAGGTTCCAGAGGACCTGCCCCCCGGGACTGTCTTGACATTTCTGGATGCCTCTGATCCTGACCTGGGCCC
CGCAGGTGAAGTGCGATATGTTCTGATGGATGGCGCCCATGGGACCTTCCGGGTGGACCTGATGACAGGGGC
GCTCATTCTGGAGAGAGAGCTGGACTTTGAGAGGCGAGCTGGGTACAATCTGAGCCTGTGGGCCAGTGATGG
TGGGAGGCCCCTAGCCCGCAGGACTCTCTGCCATGTGGAGGTGATCGTCCTGGATGTGAATGAGAATCTCCA
CCCTCCCCACTTTGCCTCCTTCGTGCACCAGGGCCAGGTGCAGGAGAACAGCCCCTCGGGAACTCAGGTGAT
TGTAGTGGCTGCCCAGGACGATGACAGTGGCTTGGATGGGGAGCTCCAGTACTTCCTGCGTGCTGGCACTGG
ACTCGCAGCCTTCAGCATCAACCAAGATACAGGAATGATTCAGACTCTGGCACCCCTGGACCGAGAATTTGC
ATCTTACTACTGGTTGACGGTATTAGCAGTGGACAGGGGTTCTGTGCCCCTCTCTTCTGTAACTGAAGTCTA
CATCGAGGTTACGGATGCCAATGACAACCCACCCCAGATGTCCCAAGCTGTGTTCTACCCCTCCATCCAGGA
GGATGCTCCCGTGGGCACCTCTGTGCTTCAACTGGATGCCTGGGACCCAGACTCCAGCTCCAAAGGGAAGCT
GACCTTCAACATCACCAGTGGGAACTACATGGGATTCTTTATGATTCACCCTGTTACAGGTCTCCTATCTAC

AGCCCAGCAGCTGGACAGAGAGAACAAGGATGAACACATCCTGGAGGTGACTGTGCTGGACAATGGGGAACC
CTCACTGAAGTCCACCTCCAGGGTGGTGGTAGGCATCTTGGACGTCAATGACAATCCACCTATATTCTCCCA
CAAGCTCTTCAATGTCCGCCTTCCAGAGAGGCTGAGCCCTGTGTCCCCTGGGCCTGTGTACAGGCTGGTGGC
TTCAGACCTGGATGAGGGTCTTAATGGCAGAGTCACCTACAGTATCGAGGACAGCTATGAGGAGGCCTTCAG
TATCGACCTGGTCACAGGTGTGGTTTCATCCAACAGCACTTTTACAGCTGGAGAGTACAACATCCTAACGAT
CAAGGCAACAGACAGTGGGCAGCCACCACTCTCAGCCAGTGTCCGGCTACACATTGAGTGGATCCCTTGGCC
CCGGCCGTCCTCCATCCCTCTGGCCTTTGATGAGACCTACTACAGCTTTACGGTCATGGAGACGGACCCTGT
GAACCACATGGTGGGGGTCATCAGCGTAGAGGGCAGACCCGGACTCTTCTGGTTCAACATCTCAGGTGGGGA
TAAGGACATGGACTTTGACATTGAGAAGACCACAGGCAGCATCGTCATTGCCAGGCCTCTTGATACCAGGAG
AAGGTCGAACTATAACTTGACTGTTGAGGTGACAGATGGGTCCCGCACCATTGCCACACAGGTCCACATCTT
CATGATTGCCAACATTAACCACCATCGGCCCCAGTTTCTGGAAACTCGTTATGAAGTCAGAGTTCCCCAGGA
CACCGTGCCAGGGGTAGAGCTCCTGCGAGTCCAGGCCATAGATCAAGACAAGGGCAAAAGCCTCATCTATAC
CATACATGGCAGCCAAGACCCAGGAAGTGCCAGCCTCTTCCAGCTGGACCCAAGCAGTGGTGTCCTGGTAAC
GGTGGGAAAATTGGACCTCGGCTCGGGGCCCTCCCAGCACACACTGACAGTCATGGTCCGAGACCAGGAAAT
ACCTATCAAGAGGAACTTCGTGTGGGTGACCATTCATGTGGAGGATGGAAACCTCCACCCACCCCGCTTCAC
TCAGCTCCATTATGAGGCAAGTGTTCCTGACACCATAGCCCCCGGCACAGAGCTGCTGCAGGTCCGAGCCAT
GGATGCTGACCGGGGAGTCAATGCTGAGGTCCACTACTCCCTCCTGAAAGGGAACAGCGAAGGTTTCTTCAA
CATCAATGCCCTGCTAGGCATCATTACTCTAGCTCAAAAGCTTGATCAGGCAAATCATGCCCCACATACTCT
GACAGTGAAGGCAGAAGATCAAGGCTCCCCACAATGGCATGACCTGGCTACAGTGATCATTCATGTCTATCC
CTCAGATAGGAGTGCCCCCATCTTTTCAAAATCTGAGTACTTTGTAGAGATCCCTGAATCAATCCCTGTTGG
TTCCCCAATCCTCCTTGTCTCTGCTATGAGCCCCTCTGAAGTTACCTATGAGTTAAGAGAGGGAAATAAGGA
TGGAGTCTTCTCTATGAACTCATATTCTGGCCTTATTTCCACCCAGAAGAAATTGGACCATGAGAAAATCTC
GTCTTACCAGCTGAAAATCCGAGGCAGCAATATGGCAGGTGCATTTACTGATGTCATGGTGGTGGTTGACAT
AATTGATGAAAATGACAATGCTCCTATGTTCTTAAAGTCAACTTTTGTGGGCCAAATTAGTGAAGCAGCTCC
ACTGTATAGCATGATCATGGATAAAAACAACAACCCCTTTGTGATTCATGCCTCTGACAGTGACAAAGAAGC
TAATTCCTTGTTGGTCTATAAAATTTTGGAGCCGGAGGCCTTGAAGTTTTTCAAAATTGATCCCAGCATGGG
AACCCTAACCATTGTATCAGAGATGGATTATGAGAGCATGCCCTCTTTCCAATTCTGTGTCTATGTCCATGA
CCAAGGAAGCCCTGTATTATTTGCACCCAGACCTGCCCAAGTCATCATTCATGTCAGAGATGTGAATGATTC
CCCTCCCAGATTCTCAGAACAGATATATGAGGTAGCAATAGTCGGGCCTATCCATCCAGGCATGGAGCTTCT
CATGGTGCGGGCCAGCGATGAAGACTCAGAAGTCAATTATAGCATCAAAACTGGCAATGCTGATGAAGCTGT
TACCATCCATCCTGTCACTGGTAGCATATCTGTGCTGAATCCTGCTTTCCTGGGACTCTCTCGGAAGCTCAC
CATCAGGGCTTCTGATGGCTTGTATCAAGACACTGCGCTGGTAAAAATTTCTTTGACCCAAGTGCTTGACAA
AAGCTTGCAGTTTGATCAGGATGTCTACTGGGCAGCTGTGAAGGAGAACTTGCAGGACAGAAAGGCACTGGT
GATTCTTGGTGCCCAGGGCAATCATTTGAATGACACCCTTTCCTACTTTCTCTTGAATGGCACAGATATGTT
TCATATGGTCCAGTCAGCAGGTGTGTTGCAGACAAGAGGTGTGGCGTTTGACCGGGAGCAGCAGGACACTCA
TGAGTTGGCAGTGGAAGTGAGGGACAATCGGACACCTCAGCGGGTGGCTCAGGGTTTGGTCAGAGTCTCTAT
TGAGGATGTCAATGACAATCCCCCCAAATTTAAGCATCTGCCCTATTACACAATCATCCAAGATGGCACAGA
GCCAGGGGATGTCCTCTTTCAGGTATCTGCCACTGATGAGGACTTGGGGACAAATGGGGCTGTTACATATGA
ATTTGCAGAAGATTACACATATTTCCGAATTGACCCCTATCTTGGGGACATATCACTCAAGAAACCCTTTGA
TTATCAAGCTTTAAATAAATATCACCTCAAAGTCATTGCTCGGGATGGAGGAACGCCATCCCTCCAGAGTGA
GGAAGAGGTACTTGTCACTGTGAGAAATAAATCCAACCCACTGTTTCAGAGTCCTTATTACAAAGTCAGAGT
ACCTGAAAATATCACCCTCTATACCCCAATTCTCCACACCCAGGCCCGGAGTCCAGAGGGACTCCGGCTCAT
CTACAACATTGTGGAGGAAGAACCCTTGATGCTGTTCACCACTGACTTCAAGACTGGTGTCCTAACAGTAAC
AGGGCCTTTGGACTATGAGTCCAAGACCAAACATGTGTTCACAGTCAGAGCCACGGATACAGCTCTGGGGTC
ATTTTCTGAAGCCACAGTGGAAGTCCTAGTGGAGGATGTCAATGATAACCCTCCCACTTTTCCCAATTGGT
CTATACCACTTCCATCTCAGAAGGCTTGCCTGCTCAGACCCCTGTGATCCAACTGTTGGCTTCTGACCAGGA
CTCAGGGCGGAACCGTGACGTCTCTTATCAGATTGTGGAGGATGGCTCAGATGTTTCCAAGTTCTTCCAGAT
CAATGGGAGCACAGGGGAGATGTCCACAGTTCAAGAACTGGATTATGAAGCCCAACAACACTTTCATGTGAA
AGTCAGGGCCATGGATAAAGGAGATCCCCCACTCACTGGTGAAACCCTTGTGGTTGTCAATGTGTCTGATAT
CAATGACAACCCCCCAGAGTTCAGACAACCTCAATATGAAGCCAATGTCAGTGAACTGGCAACCTGTGGACA
CCTGGTTCTTAAAGTCCAGGCTATTGACCCTGACAGCAGAGACACCTCCCGCCTGGAGTACCTGATTCTTTC
TGGCAATCAGGACAGGCACTTCTTCATTAACAGCTCATCGGGAATAATTTCTATGTTCAACCTTTGCAAAAA
GCACCTGGACTCTTCTTACAATTTGAGGGTAGGTGCTTCTGATGGAGTCTTCCGAGCAACTGTGCCTGTGTA
CATCAACACTACAAATGCCAACAAGTACAGCCCAGAGTTCCAGCAGCACCTTTATGAGGCAGAATTAGCAGA
GAATGCAATGGTTGGAACCAAGGTGATTGATTTGCTAGCCATAGACAAAGATAGTGGTCCCTATGGCACTAT
AGATTATACTATCATCAATAAACTAGCAAGTGAGAAGTTCTCCATAAACCCCAATGGCCAGATTGCCACTCT
GCAGAAACTGGATCGGGAAAATTCAACAGAGAGAGTCATTGCTATTAAGGTCATGGCTCGGGATGGAGGAGG
AAGAGTAGCCTTCTGCACGGTGAAGATCATCCTCACAGATGAAAATGACAACCCCCCACAGTTCAAAGCATC
TGAGTACACAGTATCCATTCAATCCAATGTCAGTAAAGACTCTCCGGTTATCCAGGTGTTGGCCTATGATGC
AGATGAAGGTCAGAACGCAGATGTCACCTACTCAGTGAACCCAGAGGACCTAGTTAAAGATGTCATTGAAAT
TAACCCAGTCACTGGTGTGGTCAAGGTGAAAGACAGCCTGGTGGGATTGGAAAATCAGACCCTTGACTTCTT
CATCAAAGCCCAAGATGGAGGCCCTCCTCACTGGAACTCTCTGGTGCCAGTACGACTTCAGGTGGTTCCTAA
AAAAGTATCCTTACCGAAATTTTCTGAACCTTTGTATACTTTCTCTGCACCTGAAGACCTTCCAGAGGGGTC
TGAAATTGGGATTGTTAAAGCAGTGGCAGCTCAAGATCCAGTCATCTACAGTCTAGTGCGGGGCACTACACC
TGAGAGCAACAAGGATGGTGTCTTCTCCCTAGACCCAGACACAGGGGTCATAAAGGTGAGGAAGCCCATGGA
CCACGAATCCACCAAATTGTACCAGATTGATGTGATGGCACATTGCCTTCAGAACACTGATGTGGTGTCCTT
GGTCTCTGTCAACATCCAAGTGGGAGACGTCAATGACAATAGGCCTGTATTTGAGGCTGATCCATATAAGGC
TGTCCTCACTGAGAATATGCCAGTGGGGACCTCAGTCATTCAAGTGACTGCCATTGACAAGGACACTGGGAG
AGATGGCCAGGTGAGCTACAGGCTGTCTGCAGACCCTGGTAGCAATGTCCATGAGCTCTTTGCCATTGACAG

TGAGAGTGGTTGGATCACCACACTCCAGGAACTTGACTGTGAGACCTGCCAGACTTATCATTTTCATGTGGT
GGCCTATGACCACGGACAGACCATCCAGCTATCCTCTCAGGCCCTGGTTCAGGTCTCCATTACAGATGAGAA
TGACAATGCTCCCCGATTTGCTTCTGAAGAGTACAGAGGATCTGTGGTTGAGAACAGTGAGCCTGGCGAACT
GGTGGCGACTCTAAAGACCCTGGATGCTGACATTTCTGAGCAGAACAGGCAGGTCACCTGCTACATCACAGA
GGGAGACCCCCTGGGCCAGTTTGGCATCAGCCAAGTTGGAGATGAGTGGAGGATTTCCTCAAGGAAGACCCT
GGACCGCGAGCATACAGCCAAGTACTTGCTCAGAGTCACAGCATCTGATGGCAAGTTCCAGGCTTCGGTCAC
TGTGGAGATCTTTGTCCTGGACGTCAATGATAACAGCCCACAGTGTTCACAGCTTCTCTATACTGGCAAGGT
TCATGAAGATGTATTTCCAGGACACTTCATTTTGAAGGTTTCTGCCACAGACTTGGACACTGATACCAATGC
TCAGATCACATATTCTCTGCATGGCCCTGGGGCGCATGAATTCAAGCTGGATCCTCATACAGGGGAGCTGAC
CACACTCACTGCCCTAGACCGAGAAAGGAAGGATGTGTTCAACCTTGTTGCCAAGGCGACGGATGGAGGTGG
CCGATCGTGCCAGGCAGACATCACCCTCCATGTGGAGGATGTGAATGACAATGCCCCGCGGTTCTTCCCCAG
CCACTGTGCTGTGGCTGTCTTCGACAACACCACAGTGAAGACCCCTGTGGCTGTAGTATTTGCCCGGGATCC
CGACCAAGGCGCCAATGCCCAGGTGGTTTACTCTCTGCCGGATTCAGCCGAAGGCCACTTTTCCATCGACGC
CACCACGGGGGTGATCCGCCTGGAAAAGCCGCTGCAGGTCAGGCCCCAGGCACCACTGGAGCTCACGGTCCG
TGCCTCTGACCTGGGCACCCCAATACCGCTGTCCACGCTGGGCACCGTCACAGTCTCGGTGGTGGGCCTAGA
AGACTACCTGCCCGTGTTCCTGAACACCGAGCACAGCGTGCAGGTGCCCGAGGACGCCCCACCTGGCACGGA
GGTGCTGCAGCTGGCCACCCTCACTCGCCCGGGCGCAGAGAAGACCGGCTACCGCGTGGTCAGCGGGAACGA
GCAAGGCAGGTTCCGCCTGGATGCTCGCACAGGGATCCTGTATGTCAACGCAAGCCTGGACTTTGAGACAAG
CCCCAAGTACTTCCTGTCCATTGAGTGCAGCCGGAAGAGCTCCTCTTCCCTCAGTGACGTGACCACAGTCAT
GGTCAACATCACTGATGTCAATGAACACCGGCCCCAATTCCCCCAAGATCCATATAGCACAAGGGTCTTAGA
GAATGCCCTTGTGGGTGACGTCATCCTCACGGTATCAGCGACTGATGAAGATGGACCCCTAAATAGTGACAT
TACCTATAGCCTCATAGGAGGGAACCAGCTTGGGCACTTCACCATTCACCCCAAAAAGGGGGAGCTACAGGT
GGCCAAGGCCCTGGACCGGGAACAGGCCTCTAGTTATTCCCTGAAGCTCCGAGCCACAGACAGTGGGCAGCC
TCCACTGCATGAGGACACAGACATCGCTATCCAAGTGGCTGATGTCAATGATAACCCACCGAGATTCTTCCA
GCTCAACTACAGCACCACTGTCCAGGAGAACTCCCCCATTGGCAGCAAAGTCCTGCAGCTGATCCTGAGTGA
CCCAGATTCTCCAGAGAATGGCCCCCCCTACTCGTTTCGAATCACCAAGGGGAACAACGGCTCTGCCTTCCG
AGTGACCCCGGATGGATGGCTGGTGACTGCTGAGGGCCTAAGCAGGAGGGCTCAGGAATGGTATCAGCTTCA
GATCCAGGCGTCAGACAGTGGCATCCCTCCCCTCTCGTCTTTGACGTCTGTCCGTGTCCATGTCACAGAGCA
GAGCCACTATGCACCTTCTGCTCTCCCACTGGAGATCTTCATCACTGTTGGAGAGGATGAGTTCCAGGGTGG
CATGGTGGGTAAGATCCATGCCACAGACCGAGACCCCCAGGACACGCTGACCTATAGCCTGGCAGAAGAGGA
GACCCTGGGCAGGCACTTCTCAGTGGGTGCGCCTGATGGCAAGATTATCGCCGCCCAGGGCCTGCCTCGTGG
CCACTACTCGTTCAACGTCACGGTCAGCGATGGGACCTTCACCACGACTGCTGGGGTCCATGTGTACGTGTG
GCATGTGGGGCAGGAGGCTCTGCAGCAGGCCATGTGGATGGGCTTCTACCAGCTCACCCCCGAGGAGCTGGT
GAGTGACCACTGGCGGAACCTGCAGAGGTTCCTCAGCCATAAGCTGGACATCAAACGGGCTAACATTCACTT
GGCCAGCCTCCAGCCTGCAGAGGCCGTGGCTGGTGTGGATGTGCTCCTGGTCTTTGAGGGGCATTCTGGAAC
CTTCTACGAGTTTCAGGAGCTAGCATCCATCATCACTCACTCAGCCAAGGAGATGGAGCATTCAGTGGGGGT
TCAGATGCGGTCAGCTATGCCCATGGTGCCCTGCCAGGGGCCAACCTGCCAGGGTCAAATCTGCCATAACAC
AGTGCATCTGGACCCCAAGGTTGGGCCCACGTACAGCACCGCCAGGCTCAGCATCCTAACCCCGCGGCACCA
CCTGCAGAGGAGCTGCTCCTGCAATGGTACTGCTACAAGGTTCAGTGGTCAGAGCTATGTGCGGTACAGGGC
CCCAGCGGCTCGGAACTGGCACATCCATTTCTATCTGAAAACACTCCAGCCACAGGCCATTCTTCTATTCAC
CAATGAAACAGCGTCCGTCTCCCTGAAGCTGGCCAGTGGAGTGCCCCAGCTGGAATACCACTGTCTGGGTGG
TTTCTATGGAAACCTTTCCTCCCAGCGCCATGTGAATGACCACGAGTGGCACTCCATCCTGGTGGAGGAGAT
GGACGCTTCCATTCGCCTGATGGTTGACAGCATGGGCAACACCTCCCTTGTGGTCCCAGAGAACTGCCGTGG
TCTGAGGCCCGAAAGGCACCTCTTGCTGGGCGGCCTCATTCTGTTGCATTCTTCCTCGAATGTCTCCCAGGG
CTTTGAAGGCTGCCTGGATGCTGTCGTGGTCAACGAAGAGGCTCTAGATCTGCTGGCCCCTGGCAAGACGGT
GGCAGGCTTGCTGGAGACACAAGCCCTCACCCAGTGCTGCCTCCACAGTGACTACTGCAGCCAGAACACATG
CCTCAATGGTGGGAAGTGCTCATGGACCCATGGGGCAGGCTATGTCTGCAAATGTCCCCCACAGTTCTCTGG
GAAGCACTGTGAACAAGGAAGGGAGAACTGTACTTTTGCACCCTGCCTGGAAGGTGGAACTTGCATCCTCTC
CCCCAAAGGAGCTTCCTGTAACTGCCCTCATCCTTACACAGGAGACAGGTGTGAAATGGAGGCGAGGGGTTG
TTCAGAAGGACACTGCCTAGTCACTCCCGAGATCCAAAGGGGGGACTGGGGGCAGCAGGAGTTACTGATCAT
CACAGTGGCCGTGGCGTTCATTATCATAAGCACTGTCGGGCTTCTCTTCTACTGCCGCCGTTGCAAGTCTCA
CAAGCCTGTGGCCATGGAGGACCCAGACCTCCTGGCCAGGAGTGTTGGTGTTGACACCCAAGCCATGCCTGC
CATCGAGCTCAACCCATTGAGTGCCAGCTCCTGCAACAACCTCAACCAACCGGAACCCAGCAAGGCCTCTGT
TCCAAATGAACTCGTCACATTTGGACCCAATTCTAAGCAACGGCCAGTGGTCTGCAGTGTGCCCCCCAGACT
CCCGCCAGCTGCGGTCCCTTCCCACTCTGACAATGAGCCTGTCATTAAGAGAACCTGGTCCAGCGAGGAGAT
GGTGTACCCTGGCGGAGCCATGGTCTGGCCCCCTACTTACTCCAGGAACGAACGCTGGGAATACCCCCACTC
CGAAGTGACTCAGGGCCCTCTGCCGCCCTCGGCTCACCGCCACTCAACCCCAGTCGTGATGCCAGAGCCTAA
TGGCCTCTATGGGGGCTTCCCCTTCCCCCTGGAGATGGAAAACAAGCGGGCACCTCTCCCACCCCGTTACAG
CAACCAGAACCTGGAAGATCTGATGCCCTCTCGGCCCCCTAGTCCCCGGGAGCGCCTGGTTGCCCCCTGTCT
CAATGAGTACACGGCCATCAGCTACTACCACTCGCAGTTCCGGCAGGGAGGGGGAGGGCCCTGCCTGGCAGA
CGGGGGCTACAAGGGGGTGGGTATGCGCCTCAGCCGAGCTGGGCCCTCTTATGCTGTCTGTGAGGTGGAGGG
GGCACCTCTTGCAGGCCAGGGCCAGCCCCGGGTGCCCCCCAACTATGAGGGCTCTGACATGGTGGAGAGTGA
TTATGGCAGCTGTGAGGAGGTCATGTTCTAGCTTCCCATTCCCAGAGCAAGGCAGGCGGGAGGCCAAGGACT
GGACTTGGCTTATTTCTTCCTGTCTCGTAGGGGGTGAGTTGAGTGTGGCTGGGAGAGTGGGAGGGAAGCCCT
CAGCCCAGGCTGTTGTCCCTTGAAATGTGCTCTTCCAATCCCCCACCTAGTCCCTGAGGGTGGAGGGAAGCT
GAGGATAGAGCTCCAGAAACAGCACTAGGGTCCCAGGAGAGGGGCATTTCTAGAGCAGTGACCCTGGAAAAC
CAGGAACAATTGACTCCTGGGGTGGGCGACAGACAGGAGGGCTCCCTGATCTGCCGGCTCTCAGTCCCCGGG
GCAAAGCCTGATTGACTGTGCTGGCTCAACTTCACCAAGATGCATTCTCATACCTGCCCACAGCTCCATTTT
GGAGGCAGGCAGGTTGGTGCCTGACAGACAACCACTACGCGGGCCGTACAGAGGAGCTCTAGAGGGCTGCGT

GGCATCCTCCTAGGGGCTGAGAGGTGAGCAGCAGGGGAGCGGGCACAGTCCCCTCTGCCCCTGCCTCAGTCG
AGCACTCACTGTGTCTTTGTCAAGTGTCTGCTCCACGTCAGGCACTGTGCTTTGCACCGGGGAGAAAATGGT
GATGGAGGGCAACAAGGACTCCGAGGAGCACCACCAGGCCTCGGGCCCCAGAGGTCCCGCTCCTCAGCCTAC
ACGCAGAGGAACGGGCCCACCTCAGAGTCACACCACTGGCTGCCAGTCAGGGCCTGCCAGGAGTCTACACAG
CTCTGAACCTTCTTTGTTAAAGAATTCAGACCTCATGGAACTCTGGGTTCTTCATCCCAAGTTTCCCAGGCA
CTTTTGGCCAAAGGAAGGAAGGAACTAATTCTTCATTTTAAAAATTCTTAGGCACTTTTTGACCTTGCTGTC
TGGATGAGTTTCCTCAATGGGATTTTTCTTCCCTAGACACAAGGAAGTCTGAACTCCTATTTAGGGCCGGTT
GGAAGCAGGGAGCTGGACCGCAGTGTCCAGGCTGGACACCTGCCATTGCCTCCTCTCCACTGCAGACGCCTG
CCCATCAAGTATTACCTGCAGCGACTCAACCCTATGCATGGAGGGTCAATGTGGGCACATGTCTACACATGT
GGGTGCCCATGGATAGTACGTGTGTACACATGTGTAGAGTGTATGTAGCCAGGAGTGGTGGGGACCAGAAGC
CTCTGTGGCCTTTGGTGACCTCACCACTCCCTCCCACCCAGTCCCTCCCTCTGGTCCACTGCCTTTTCATAT
GTGTTGTTTCTGGAGACAGAAGTCAAAAGGAAGAGCAGTGGAGCCTTGCCCACAGGGCTGCTGCTTCATGCG
AGAGGGAGATGTGTGGGCGAGAGCCAATTTGTGTGAGTGGTTTGTGGCTGTGTGTGTGACTGTGAGTGTGAG
TGACAGATACATAGTTTCATTGGTCATTTTTTTTTTTAACAATAAAGTATCTTTTTTTACTGTT

The disclosed NOV2 nucleic acid sequence, localized to the q33 region of human chromosome 5, has 14536 of 14536 bases (100%) identical to a protocadherin Fat 2 (FAT2) cadherin related tumor suppressor (GENBANK-ID: AF231022) (E = 0.0).

A NOV2 polypeptide (SEQ ID NO:4) encoded by SEQ ID NO:3 has 4349 amino acid residues and is presented using the one-letter code in Table 2B. Signal P, Psort and/or Hydropathy results predict that NOV2 does not contain a signal peptide and is likely to be localized in the plasma membrane, and is a Type Ia membrane protein.

Table 2B. Encoded NOV2 protein sequence (SEQ ID NO:4).

MTIALLGFAIFLLHCATCEKPLEGILSSSAWHFTHSHYNATIYENSSPKTYVESFEKMGIYLAEPQWAVRYR
IISGDVANVFKTEEYVVGNFCFLRIRTKSSNTALLNREVRDSYTLIIQATEKTLELEALTRVVVHILDQNDL
KPLFSPPSYRVTISEDMPLKSPICKVTATDADLGQNAEFYYAFNTRSEMFAIHPTSGVVTVAGKLNVTWRGK
HELQVLAVDRMRKISEGNGFGSLAALVVHVEPALRKPPAIASVVVTPPDSNDGTTYATVLVDANSSGAEVES
VEVVGGDPGKHFKAIKSYARSNEFSLVSVKDINWMEYLHGFNLSLQARSGSGPYFYSQIRGFHLPPSKLSSL
KFEKAVYRVQLSEFSPPGSRVVMVRVTPAFPNLQYVLKPSSENVGFKLNARTGLITTTKLMDFHDRAHYQLH
IRTSPGQASTVVVIDIVDCNNHAPLFNRSSYDGTLDENIPPGTSVLAVTATDRDHGENGYVTYSIAGPKALP
FSIDPYLGIISTSKPMDYELMKRIYTFRVRASDWGSPFRREKEVSIFLQLRNLNDNQPMFEEVNCTGSIRQD
WPVGKSIMTMSAIDVDELQNLKYEIVSGNELEYFDLNHFSGVISLKRPFINLTAGQPTSYSLKITASDGKNY
ASPTTLNITVVKDPHFEVPVTCDKTGVLTQFTKTILHFIGLQNQESSDEEFTSLSTYQINHYTPQFEDHFPQ
SIDVLESVPINTPLARLAATDPDAGFNGKLVYVIADGNEEGCFDIELETGLLTVAAPLDYEATNFYILNVTV
YDLGTPQKSSWKLLTVNVKDWNDNAPRFPPGGYQLTISEDTEVGTTIAELTTKDADSEDNGRVRYTLLSPTE
KFSLHPLTGELVVTGHLDRESEPRYILKVEARDQPSKGHQLPSVTDLIITLEDVNDNSPQCITEHNRLKVPE
DLPPGTVLTFLDASDPDLGPAGEVRYVLMDGAHGTFRVDLMTGALILERELDFERRAGYNLSLWASDGGRPL
ARRTLCHVEVIVLDVNENLHPPHFASFVHQGQVQENSPSGTQVIVVAAQDDDSGLDGELQYFLRAGTGLAAF
SINQDTGMIQTLAPLDREFASYYWLTVLAVDRGSVPLSSVTEVYIEVTDANDNPPQMSQAVFYPSIQEDAPV
GTSVLQLDAWDPDSSSKGKLTFNITSGNYMGFFMIHPVTGLLSTAQQLDRENKDEHILEVTVLDNGEPSLKS
TSRVVVGILDVNDNPPIFSHKLFNVRLPERLSPVSPGPVYRLVASDLDEGLNGRVTYSIEDSYEEAFSIDLV
TGVVSSNSTFTAGEYNILTIKATDSGQPPLSASVRLHIEWIPWPRPSSIPLAFDETYYSFTVMETDPVNHMV
GVISVEGRPGLFWFNISGGDKDMDFDIEKTTGSIVIARPLDTRRRSNYNLTVEVTDGSRTIATQVHIFMIAN
INHHRPQFLETRYEVRVPQDTVPGVELLRVQAIDQDKGKSLIYTIHGSQDPGSASLFQLDPSSGVLVTVGKL
DLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDGNLHPPRFTQLHYEASVPDTIAPGTELLQVRAMDADR
GVNAEVHYSLLKGNSEGFFNINALLGIITLAQKLDQANHAPHTLTVKAEDQGSPQWHDLATVIIHVYPSDRS
APIFSKSEYFVEIPESIPVGSPILLVSAMSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISSYQL
KIRGSNMAGAFTDVMVVVDIIDENDNAPMFLKSTFVGQISEAAPLYSMIMDKNNNPFVIHASDSDKEANSLL
VYKILEPEALKFFKIDPSMGTLTIVSEMDYESMPSFQFCVYVHDQGSPVLFAPRPAQVIIHVRDVNDSPPRF
SEQIYEVAIVGPIHPGMELLMVRASDEDSEVNYSIKTGNADEAVTIHPVTGSISVLNPAFLGLSRKLTIRAS
DGLYQDTALVKISLTQVLDKSLQFDQDVYWAAVKENLQDRKALVILGAQGNHLNDTLSYFLLNGTDMFHMVQ
SAGVLQTRGVAFDREQQDTHELAVEVRDNRTPQRVAQGLVRVSIEDVNDNPPKFKHLPYYTIIQDGTEPGDV
LFQVSATDEDLGTNGAVTYEFAEDYTYFRIDPYLGDISLKKPFDYQALNKYHLKVIARDGGTPSLQSEEEVL
VTVRNKSNPLFQSPYYKVRVPENITLYTPILHTQARSPEGLRLIYNIVEEEPLMLFTTDFKTGVLTVTGPLD
YESKTKHVFTVRATDTALGSFSEATVEVLVEDVNDNPPTFSQLVYTTSISEGLPAQTPVIQLLASDQDSGRN
RDVSYQIVEDGSDVSKFFQINGSTGEMSTVQELDYEAQQHFHVKVRAMDKGDPPLTGETLVVVNVSDINDNP

PEFRQPQYEANVSELATCGHLVLKVQAIDPDSRDTSRLEYLILSGNQDRHFFINSSSGIISMFNLCKKHLDS
SYNLRVGASDGVFRATVPVYINTTNANKYSPEFQQHLYEAELAENAMVGTKVIDLLAIDKDSGPYGTIDYTI
INKLASEKFSINPNGQIATLQKLDRENSTERVIAIKVMARDGGGRVAFCTVKIILTDENDNPPQFKASEYTV
SIQSNVSKDSPVIQVLAYDADEGQNADVTYSVNPEDLVKDVIEINPVTGVVKVKDSLVGLENQTLDFFIKAQ
DGGPPHWNSLVPVRLQVVPKKVSLPKFSEPLYTFSAPEDLPEGSEIGIVKAVAAQDPVIYSLVRGTTPESNK
DGVFSLDPDTGVIKVRKPMDHESTKLYQIDVMAHCLQNTDVVSLVSVNIQVGDVNDNRPVFEADPYKAVLTE
NMPVGTSVIQVTAIDKDTGRDGQVSYRLSADPGSNVHELFAIDSESGWITTLQELDCETCQTYHFHVVAYDH
GQTIQLSSQALVQVSITDENDNAPRFASEEYRGSVVENSEPGELVATLKTLDADISEQNRQVTCYITEGDPL
GQFGISQVGDEWRISSRKTLDREHTAKYLLRVTASDGKFQASVTVEIFVLDVNDNSPQCSQLLYTGKVHEDV
FPGHFILKVSATDLDTDTNAQITYSLHGPGAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATDGGGRSCQ
ADITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSAEGHFSIDATTGV
IRLEKPLQVRPQAPLELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPVFLNTEHSVQVPEDAPPGTEVLQL
ATLTRPGAEKTGYRVVSGNEQGRFRLDARTGILYVNASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNIT
DVNEHRPQFPQDPYSTRVLENALVGDVILTVSATDEDGPLNSDITYSLIGGNQLGHFTIHPKKGELQVAKAL
DREQASSYSLKLRATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSP
ENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTSVRVHVTEQSHYA
PSALPLEIFITVGEDEFQQGGMVGKIHATDRDPQDTLTYSLAEEETLGRHFSVGAPDGKIIAAQGLPRGHYSF
NVTVSDGTFTTTAGVHVYVWHVGQEALQQAMWMGFYQLTPEELVSDHWRNLQRFLSHKLDIKRANIHLASLQ
PAEAVAGVDVLLVFEGHSGTFYEFQELASIITHSAKEMEHSVGVQMRSAMPMVPCQGPTCQGQICHNTVHLD
PKVGPTYSTARLSILTPRHHLQRSCSCNGTATRFSGQSYVRYRAPAARNWHIHFYLKTLQPQAILLFTNETA
SVSLKLASGVPQLEYHCLGGFYGNLSSQRHVNDHEWHSILVEEMDASIRLMVDSMGNTSLVVPENCRGLRPE
RHLLLGGLILLHSSSNVSQGFEGCLDAVVVNEEALDLLAPGKTVAGLLETQALTQCCLHSDYCSQNTCLNGG
KCSWTHGAGYVCKCPPQFSGKHCEQGRENCTFAPCLEGGTCILSPKGASCNCPHPYTGDRCEMEARGCSEGH
CLVTPEIQRGDWGQQELLIITVAVAFIIISTVGLLFYCRRCKSHKPVAMEDPDLLARSVGVDTQAMPAIELN
PLSASSCNNLNQPEPSKASVPNELVTFGPNSKQRPVVCSVPPRLPPAAVPSHSDNEPVIKRTWSSEEMVYPG
GAMVWPPTYSRNERWEYPHSEVTQGPLPPSAHRHSTPVVMPEPNGLYGGFPFPLEMENKRAPLPPRYSNQNL
EDLMPSRPPSPRERLVAPCLNEYTAISYYHSQFRQGGGGPCLADGGYKGVGMRLSRAGPSYAVCEVEGAPLA
GQGQPRVPPNYEGSDMVESDYGSCEEVMF

The disclosed NOV2 amino acid sequence has 4349 of 4349 amino acid residues (100%) identical to, and 4349 of 4349 amino acid residues (100%) similar to, the 4349 amino acid residue Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor protein from human protocadherin Fat 2 (FAT2) cadherin related tumor suppressor (GENBANK-ID: AF231022) (E = 0.0).

TaqMan data for NOV2 is displayed below in Example 1, and SAGE data is shown below in Example 2. The TaqMan data shows overexpression of NOV2 in ovarian cancer cell lines, breast and lung cancers and high expression in cerebellum. Sage analysis agrees for Cerebellum and weaker for Ovaries.

NOV2 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 2C.

Table 2C. BLAST results for NOV2

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect

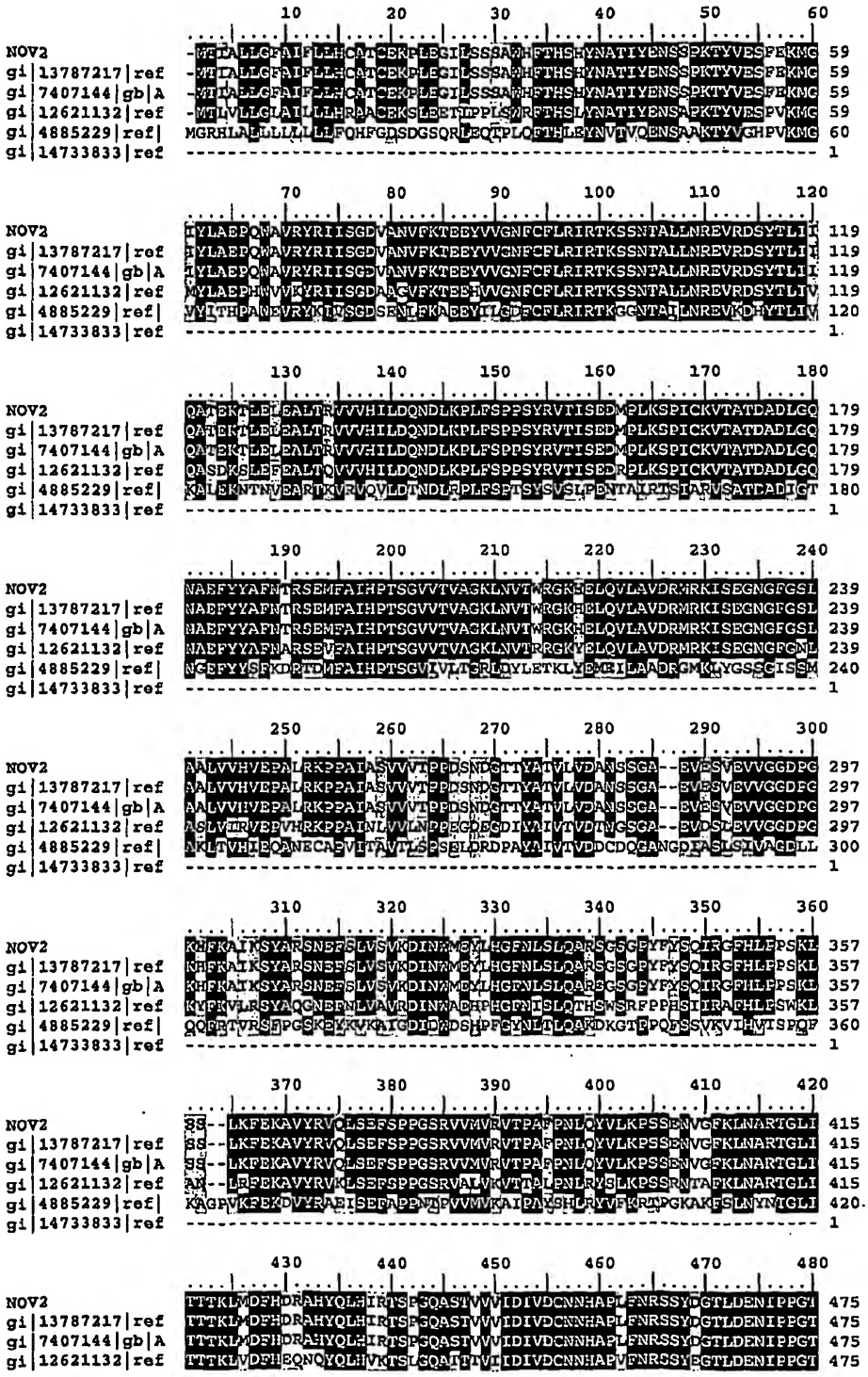
gi|13787217|ref|NP_001438.1|	FAT tumor suppressor 2 precursor; multiple epidermal growth factor-like domains 1; cadherin family member 8; FAT tumor suppressor (Drosophila) homolog 2; protocadherin FAT2 [Homo sapiens]	4349	4305/4349 (98%)	4306/4349 (98%)	0.0
gi|7407144|gb|AAF61928.1|AF231022_1 (AF231022)	protocadherin Fat 2 [Homo sapiens]	4349	4307/4349 (99%)	4307/4349 (99%)	0.0
gi|12621132|ref|NP_075243.1|	MEGF1 [Rattus norvegicus]	4351	3524/4351 (80%)	3878/4351 (88%)	0.0
gi|4885229|ref|NP_005236.1|	FAT tumor suppressor precursor; homolog of Drosophila tumor suppressor Fat precursor; cadherin-related tumor suppressor homolog precursor; homolog of Drosophila Fat protein precursor; homolog of Drosophila Fat protein; cadherin family member 7 precursor	4590	1828/4089 (44%)	2623/4089 (63%)	0.0
gi|14733833|ref|XP_041971.1|	FAT tumor suppressor 2 precursor [Homo sapiens]	2991	2963/2991 (99%)	2963/2991 (99%)	0.0

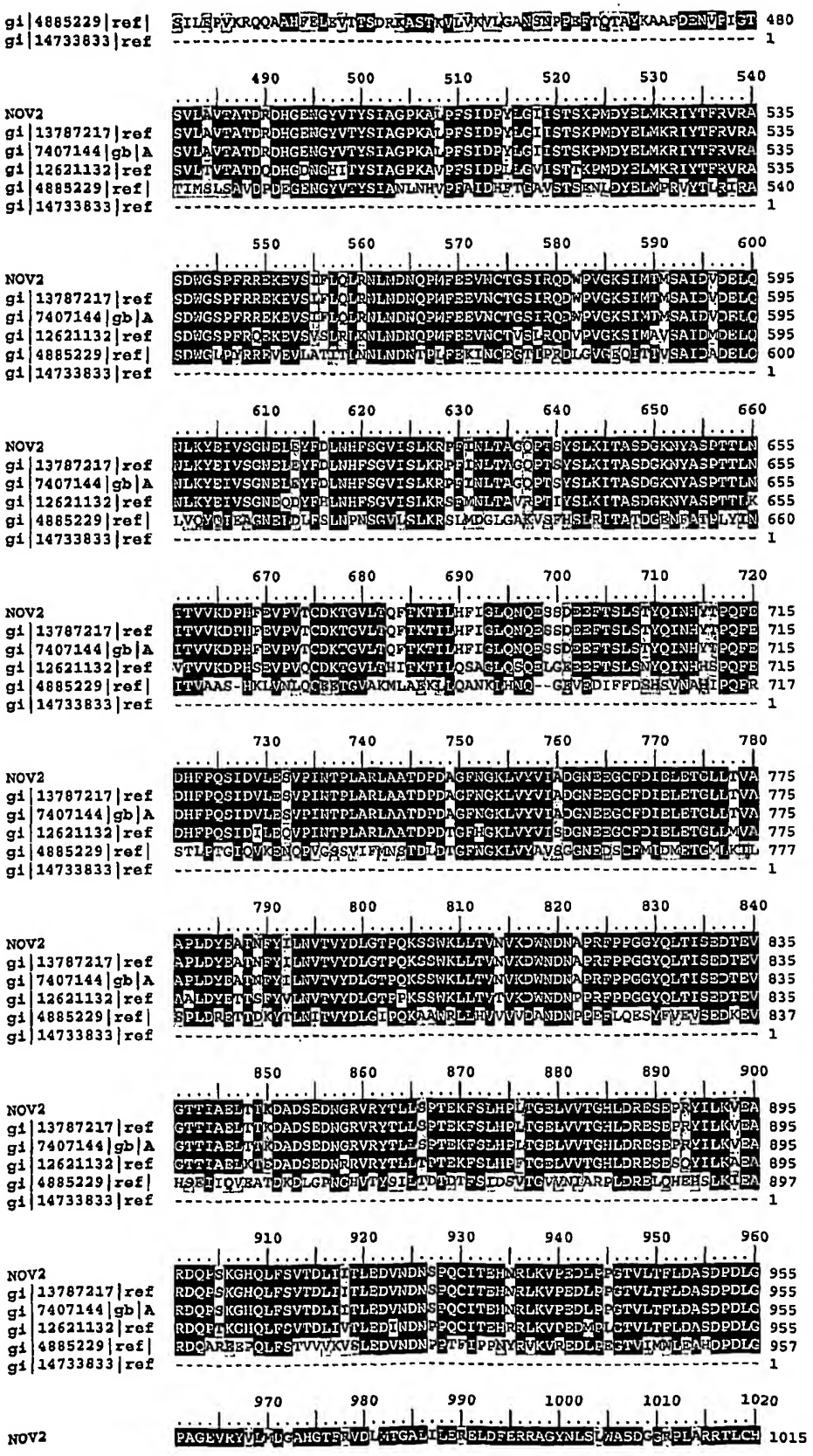
The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 2D.

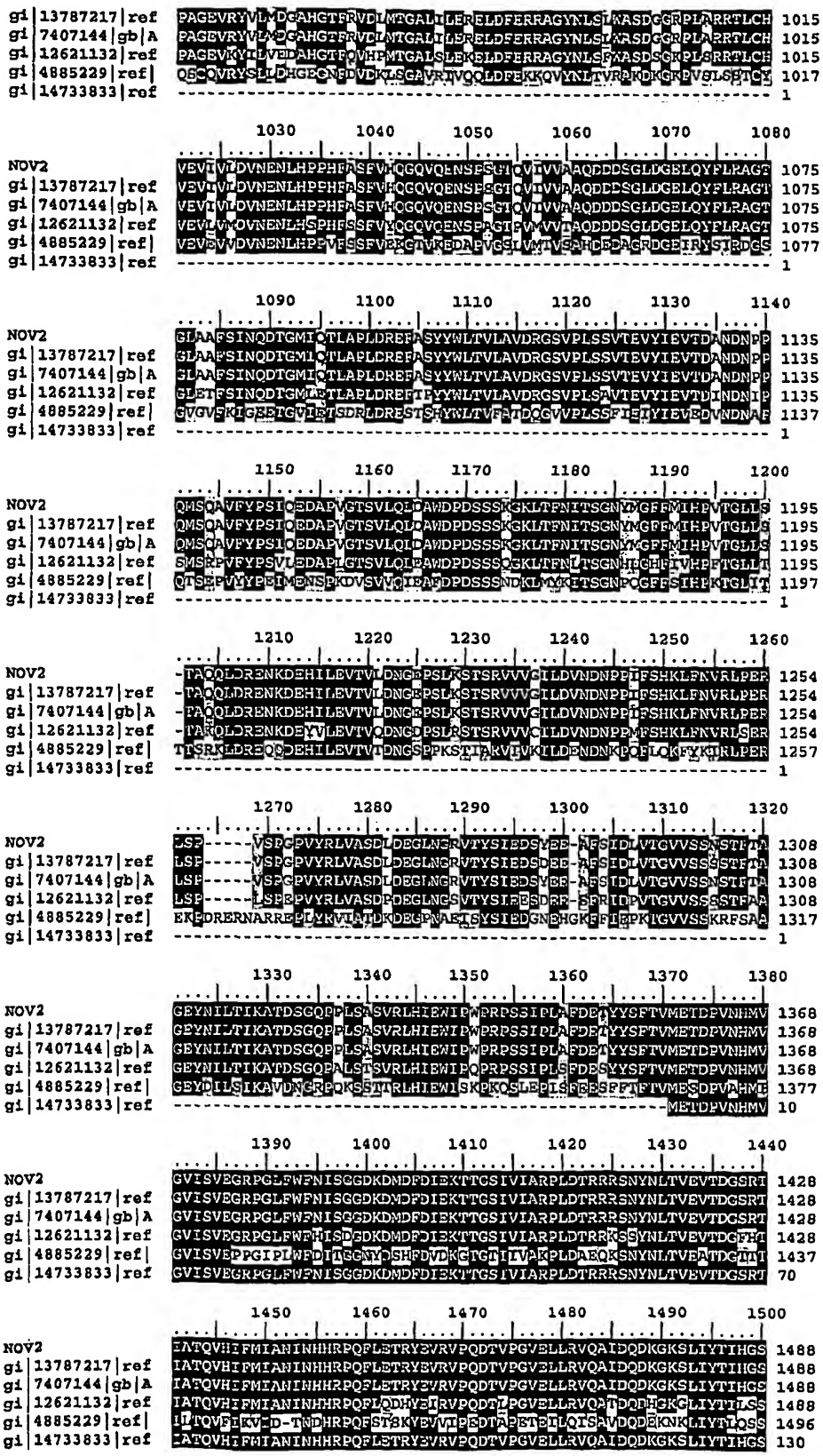
Table 2D. ClustalW Analysis of NOV2

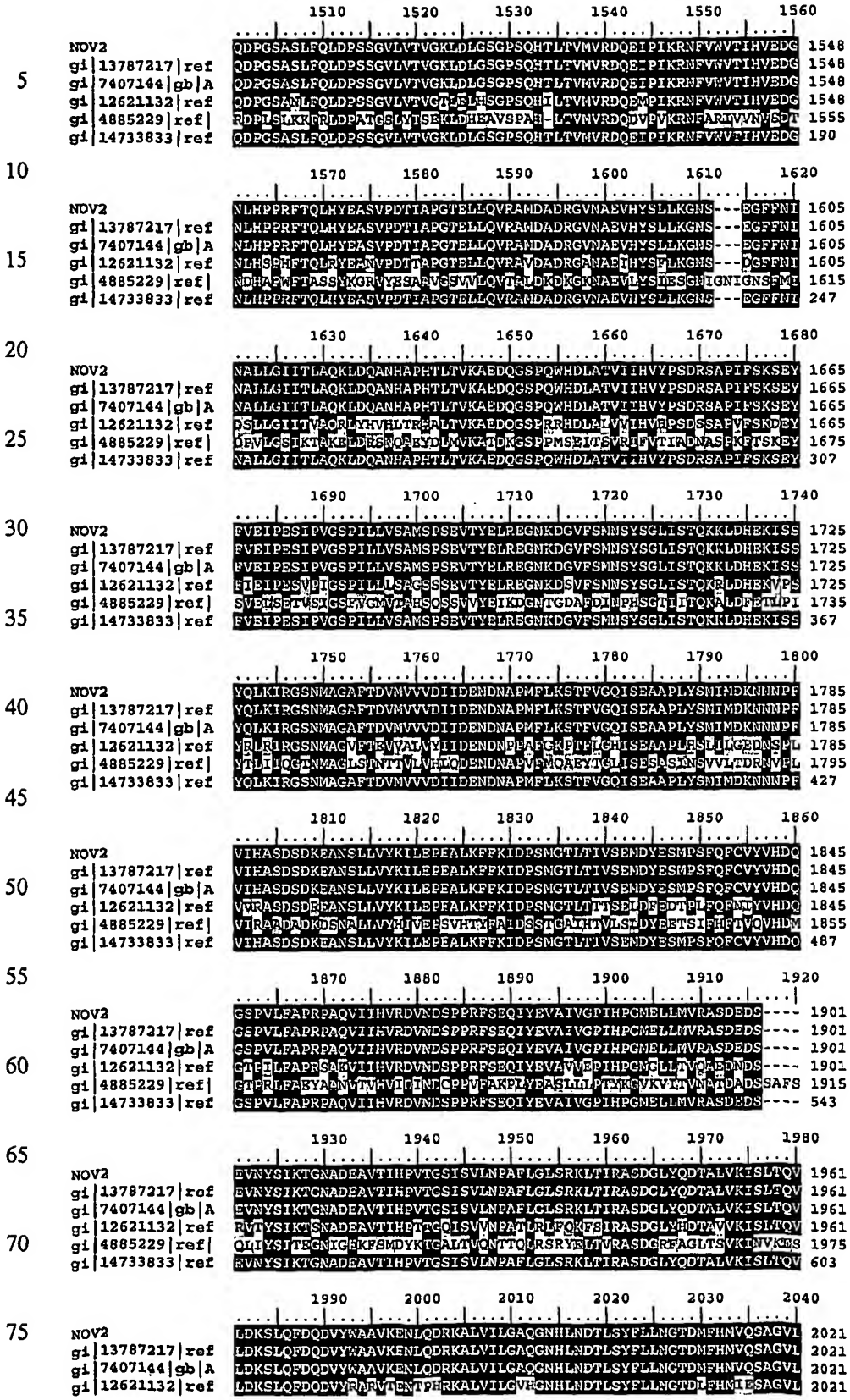
1) NOV2 (SEQ ID NO:4)
2) gi|13787217|ref|NP_001438.1| FAT tumor suppressor 2 precursor (SEQ ID NO:18)

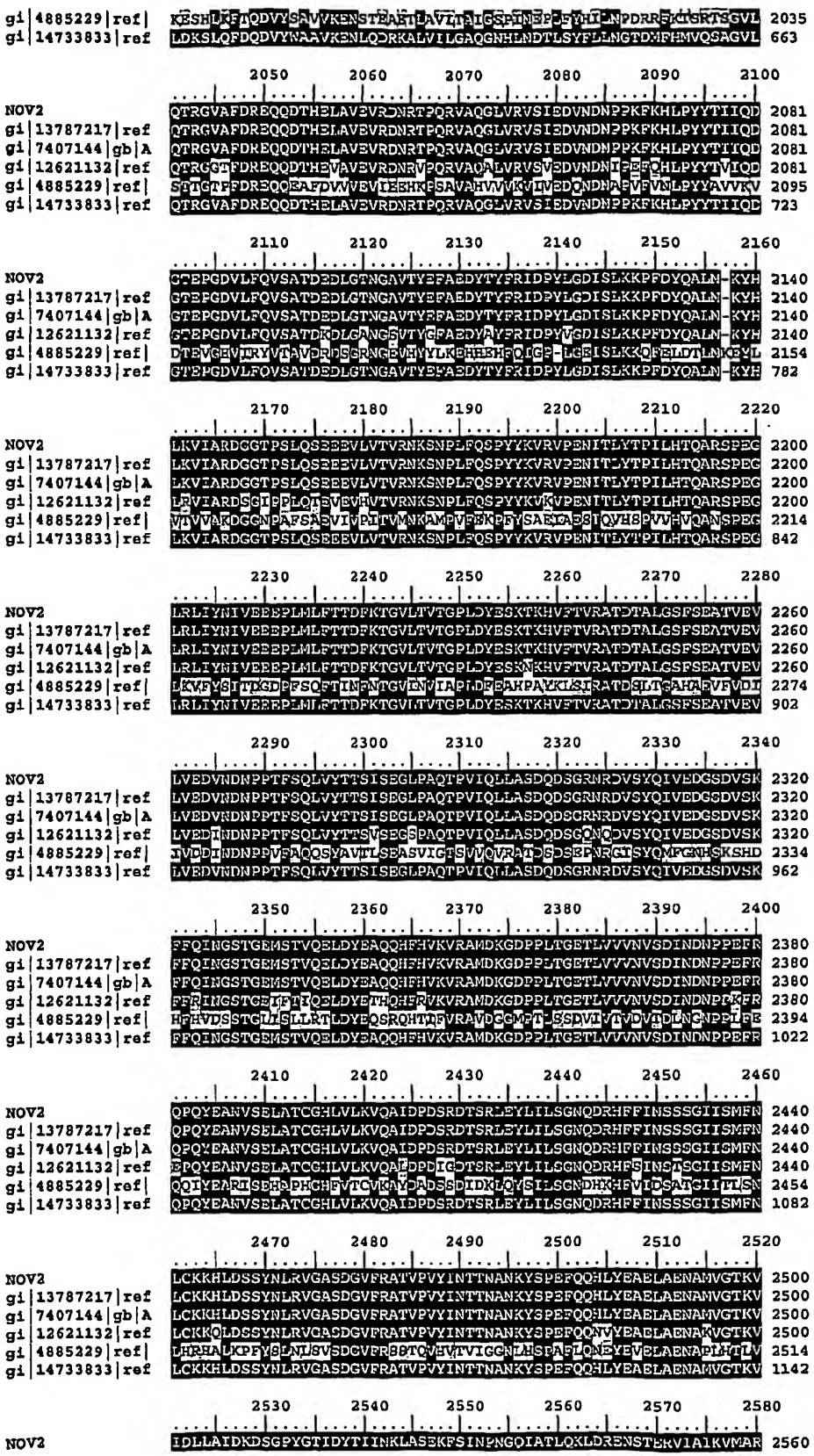
2) gi|7407144|gb|AAF61928.1|AF231022_1 (AF231022) protocadherin Fat 2 [Homo sapiens] (SEQ ID NO:19)
3) gi|12621132|ref|NP_075243.1| MEGF1 [Rattus norvegicus] (SEQ ID NO:20)
4) gi|4885229|ref|NP_005236.1| FAT tumor suppressor precursor (SEQ ID NO:21)
5) gi|14733833|ref|XP_041971.1| FAT tumor suppressor 2 precursor [Homo sapiens] (SEQ ID NO:22)

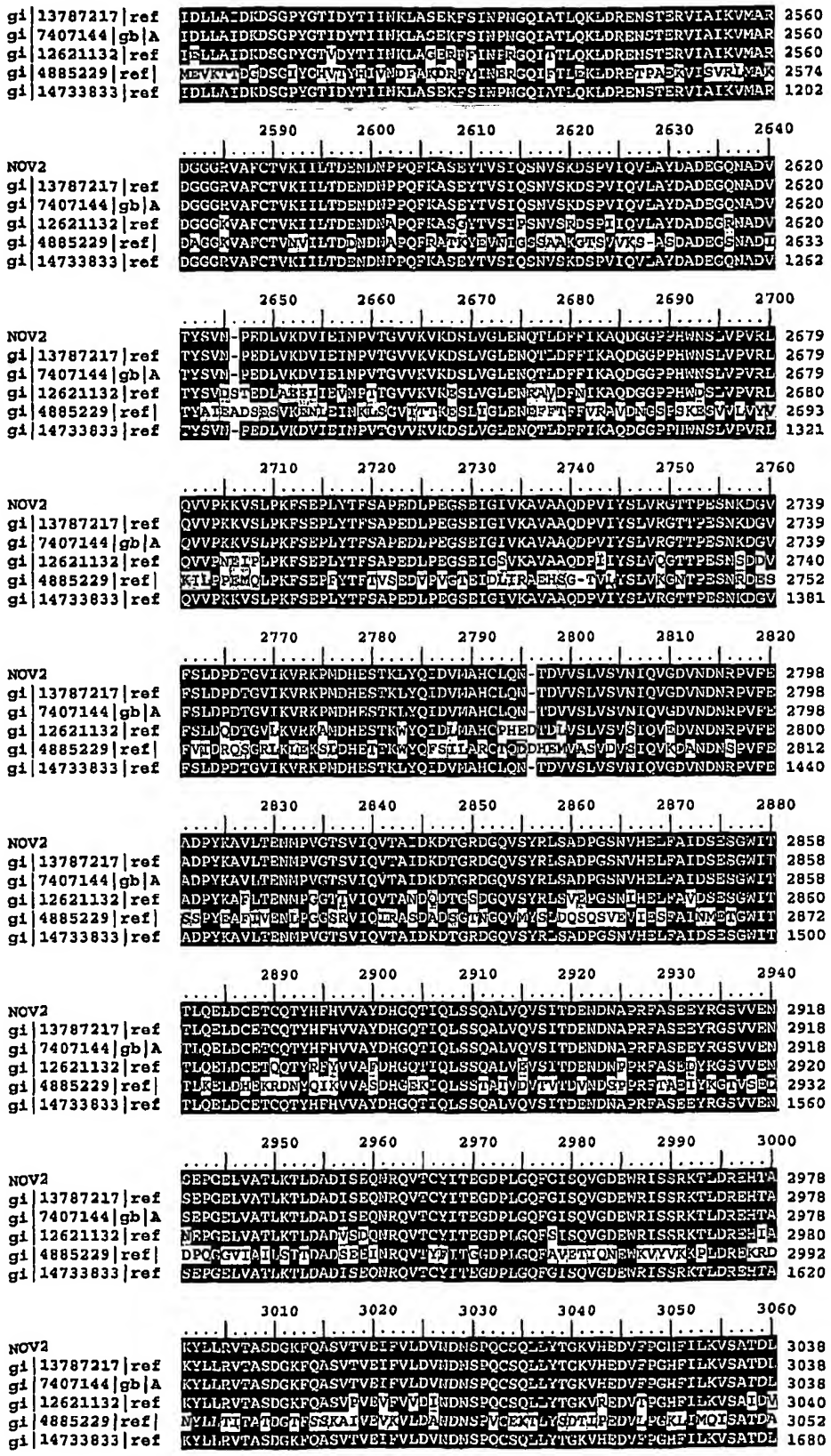


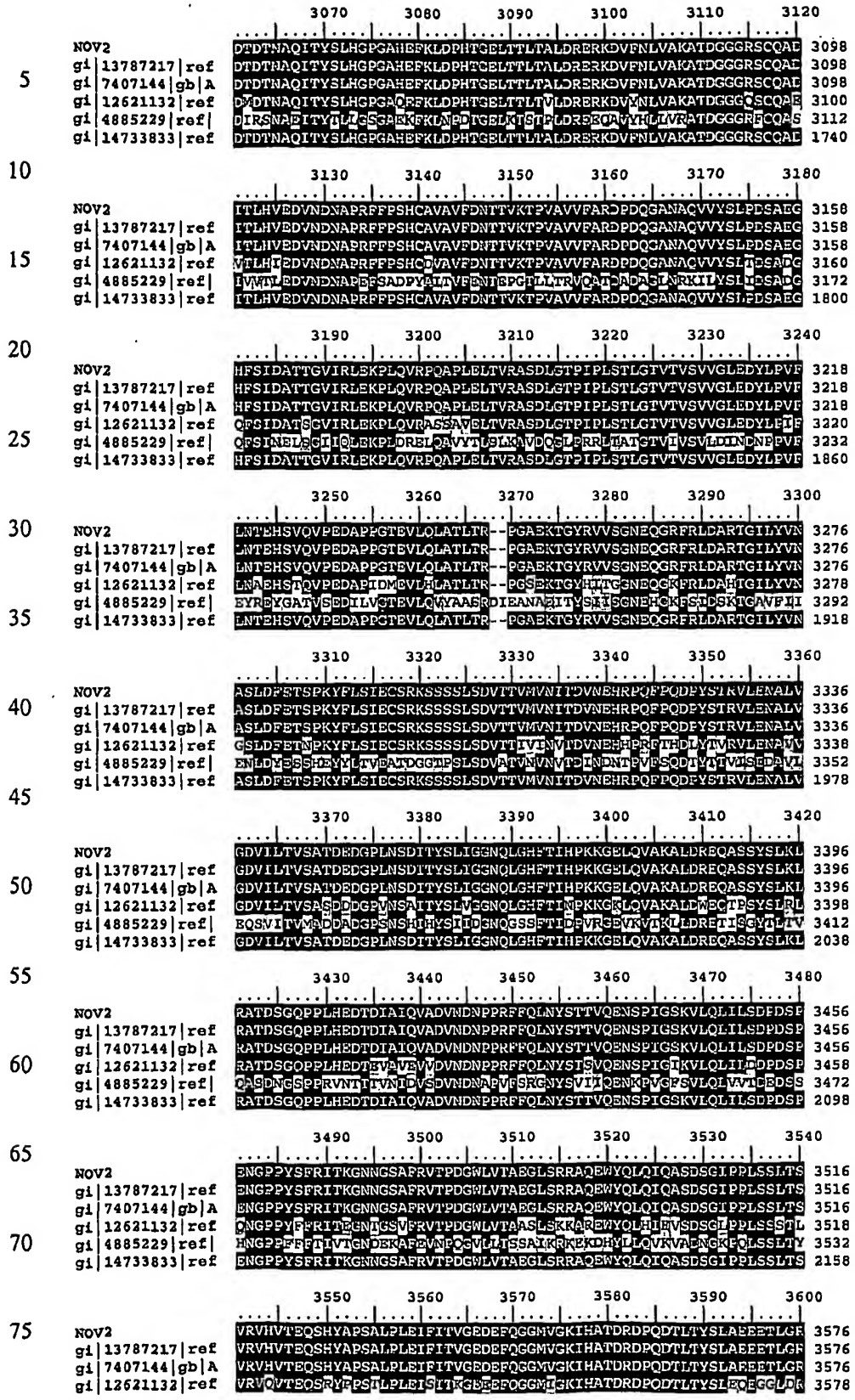


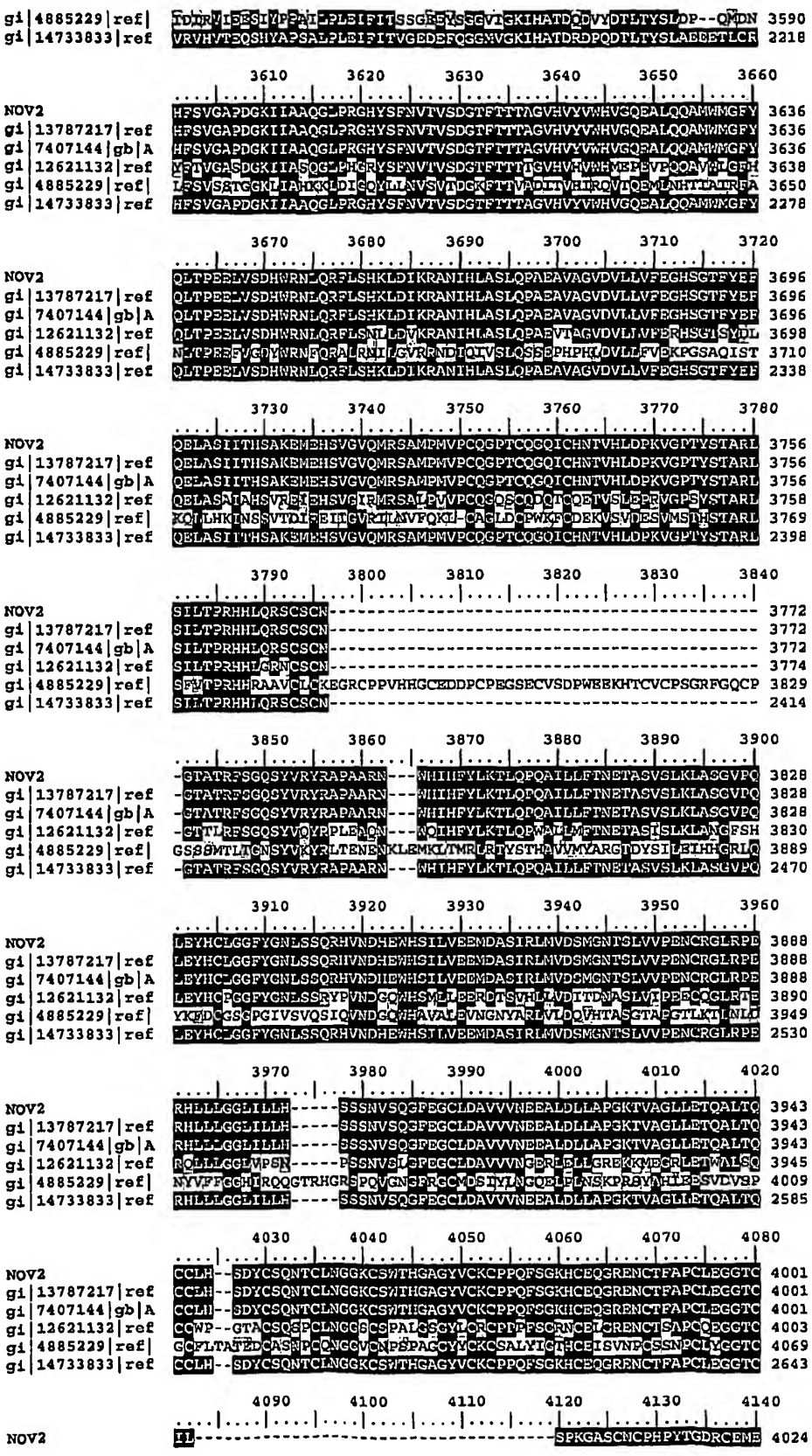


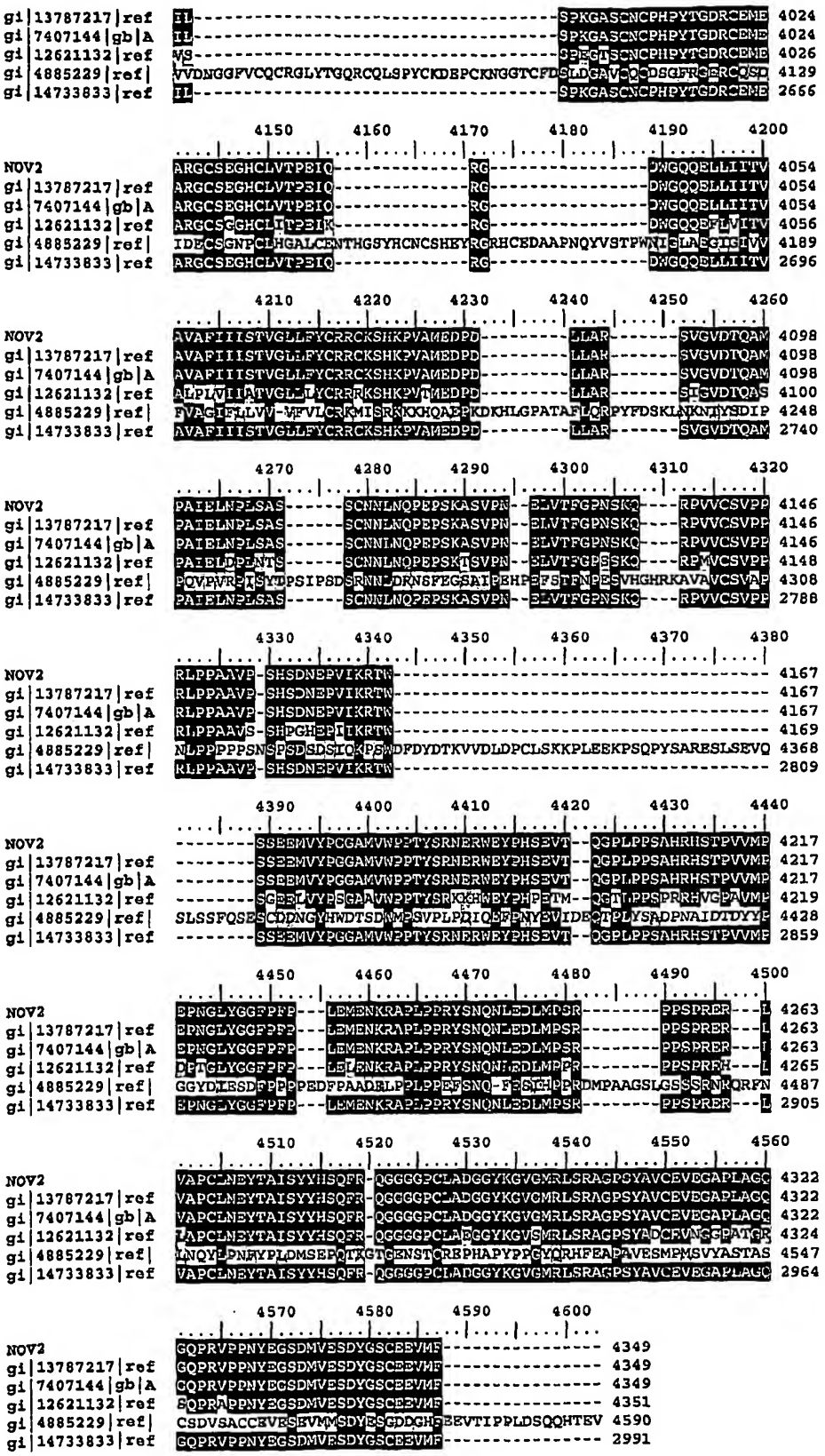

1510 1520 1530 1540 1550 1560
NOV2 QDPGSASLFQLDPSSGVLVTVGKLDLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDG 1548
gi|13787217|ref QDPGSASLFQLDPSSGVLVTVGKLDLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDG 1548
gi|7407144|gb|A QDPGSASLFQLDPSSGVLVTVGKLDLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDG 1548
gi|12621132|ref QDPGSANLFQLDPSSGVLVTVGTLDLHSGPSQHILTVMVRDQEMPIKRNFVWVTIHVEDG 1548
gi|4885229|ref| RDPLSLKKFRLDPATGSLYISEKLDHEAVSPAH-LTVMVRDQDVPVKRNFARIVVNVSDT 1555
gi|14733833|ref QDPGSASLFQLDPSSGVLVTVGKLDLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDG 190
1570 1580 1590 1600 1610 1620
NOV2 NLHPPRFTQLHYEASVPDTIAPGTELLQVRANDADRGVNAEVHYSLLKGNS---EGFFNI 1605
gi|13787217|ref NLHPPRFTQLHYEASVPDTIAPGTELLQVRANDADRGVNAEVHYSLLKGNS---EGFFNI 1605
gi|7407144|gb|A NLHPPRFTQLHYEASVPDTIAPGTELLQVRANDADRGVNAEVHYSLLKGNS---EGFFNI 1605
gi|12621132|ref NLHSPHFTQLRYEANVPDTIAPGTELLQVRAVDADRGANAEIHYSFLKGNS---QGFFNI 1605
gi|4885229|ref| NDHAPWFTASSYKGRVYESAAVSSVVLQVTALDKDKGKNAEVLYSIESGNIGNIGNSFMI 1615
gi|14733833|ref NLHPPRFTQLHYEASVPDTIAPGTELLQVRANDADRGVNAEVHYSLLKGNS---EGFFHI 247
1630 1640 1650 1660 1670 1680
NOV2 NALLGIITLAQKLDQANHAPHTLTVKAEDQGSPQWHDLATVIIHVYPSDRSAPIFSKSEY 1665
gi|13787217|ref NALLGIITLAQKLDQANHAPHTLTVKAEDQGSPQWHDLATVIIHVYPSDRSAPIFSKSEY 1665
gi|7407144|gb|A NALLGIITLAQKLDQANHAPHTLTVKAEDQGSPQWHDLATVIIHVYPSDRSAPIFSKSEY 1665
gi|12621132|ref DSLLGIITVAQRLYHVHLTRHALTVKAEDQGSPRRHDLALVVIHVHPSDSSAPVFSKDEY 1665
gi|4885229|ref| DPVLGSIKTAKELDESNQAEYDLMVKATDKGSPPMSEITSVRIFVTIRDVNASPKFTSKEY 1675
gi|14733833|ref NALLGIITLAQKLDQANHAPHTLTVKAEDQGSPQWHDLATVIIHVYPSDRSAPIFSKSEY 307
1690 1700 1710 1720 1730 1740
NOV2 FVEIPESIPVGSPILLVSAMSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISS 1725
gi|13787217|ref FVEIPESIPVGSPILLVSAMSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISS 1725
gi|7407144|gb|A FVEIPESIPVGSPILLVSAMSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISS 1725
gi|12621132|ref FIEIPESVPIGSPILLLSAGSSSEVTYELREGNKDSVFSMNSYSGLISTQKRLDHEKVFS 1725
gi|4885229|ref| SVELSETVSIGSFVGMVTAHSQSSVVYEIKDGNTGDAFDINPHSGTIITQKALDFETLPI 1735
gi|14733833|ref FVEIPESIPVGSPILLVSAMSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISS 367
1750 1760 1770 1780 1790 1800
NOV2 YQLKIRGSNMAGAFTDVMVVVDIIDENDNAPMFLKSTFVGQISEAAPLYSNIMDKNNNPF 1785
gi|13787217|ref YQLKIRGSNMAGAFTDVMVVVDIIDENDNAPMFLKSTFVGQISEAAPLYSNIMDKNNNPF 1785
gi|7407144|gb|A YQLKIRGSNMAGAFTDVMVVVDIIDENDNAPMFLKSTFVGQISEAAPLYSMIMDKNNNPF 1785
gi|12621132|ref YRLRIRGSNMAGVFTKVALVYIIDENDNPPAFGKPTFHGHISEAAPLFSLILGEDNSPL 1785
gi|4885229|ref| YTLIIQGTNMAGLSTNTTVLVHLQDENDNAPVFMQAEYTGLISESASINSVVLTDRNVPL 1795
gi|14733833|ref YQLKIRGSNMAGAFTDVMVVVDIIDENDNAPMFLKSTFVGQISEAAPLYSNIMDKNNNPF 427
1810 1820 1830 1840 1850 1860
NOV2 VIHASDSDKEANSLLVYKILEPEALKFFKIDPSMGTLTIVSENDYESMPSFQFCVYVHDQ 1845
gi|13787217|ref VIHASDSDKEANSLLVYKILEPEALKFFKIDPSMGTLTIVSENDYESMPSFQFCVYVHDQ 1845
gi|7407144|gb|A VIHASDSDKEANSLLVYKILEPEALKFFKIDPSMGTLTIVSENDYESMPSFQFCVYVHDQ 1845
gi|12621132|ref VVRASDSDEEANSLLVYKILEPEALKFFKIDPSMGTLTTSELDFEDTPLFQFNLYVHDQ 1845
gi|4885229|ref| VIRAADADKDSNALLVYHIVEPSVHTYFAIDSSTGAIHTVLSLDYEETSIFHFLVQVHDM 1855
gi|14733833|ref VIHASDSDKEANSLLVYKILEPEALKFFKIDPSMGTLTIVSENDYESMPSFQFCVYVHDQ 487
1870 1880 1890 1900 1910 1920
NOV2 GSPVLFAPRPAQVIIHVRDVNDSPPRFSEQIYEVAIVGPIHPGMELLMVRASDEDS---- 1901
gi|13787217|ref GSPVLFAPRPAQVIIHVRDVNDSPPRFSEQIYEVAIVGPIHPGMELLMVRASDEDS---- 1901
gi|7407144|gb|A GSPVLFAPRPAQVIIHVRDVNDSPPRFSEQIYEVAIVGPIHPGMELLMVRASDEDS---- 1901
gi|12621132|ref GTGILFAPRSAKVIIHVRDVNDSPPRFSEQIYEVAVVEPIHPGNGLLTVQAEDNDS---- 1901
gi|4885229|ref| GTPRLFAEYAANVTVHVIQINDCPPVFAKPLYEASLLLPIYKGVKVITVNATDADSSAFS 1915
gi|14733833|ref GSPVLFAPRPAQVIIHVRDVNDSPPKFSEQIYEVAIVGPIHPGMELLMVRASDEDS---- 543
1930 1940 1950 1960 1970 1980
NOV2 EVNYSIKTGNADEAVTIHPVTGSISVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQV 1961
gi|13787217|ref EVNYSIKTGNADEAVTIHPVTGSISVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQV 1961
gi|7407144|gb|A EVNYSIKTGNADEAVTIHPVTGSISVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQV 1961
gi|12621132|ref RVTYSIKTSNADEAVTIHPITGQISVVNPATLRLFQKFSIRASDGLYHDTAVVKISLTQV 1961
gi|4885229|ref| QLIYSITEGNIGEKFSMDYKTGALTVQNTTQLRSRYELTVRASDGRFAGLTSVKINVKDS 1975
gi|14733833|ref EVNYSIKTGNADEAVTIHPVTGSISVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQV 603
1990 2000 2010 2020 2030 2040
NOV2 LDKSLQFDQDVYWAAVKENLQDRKALVILGAQGNHLNDTLSYFLLNGTDMFHMVQSAGVL 2021
gi|13787217|ref LDKSLQFDQDVYWAAVKENLQDRKALVILGAQGNHLNDTLSYFLLNGTDMFHMVQSAGVL 2021
gi|7407144|gb|A LDKSLQFDQDVYWAAVKENLQDRKALVILGAQGNHLNDTLSYFLLNGTDMFHMVQSAGVL 2021
gi|12621132|ref LDKSLQFDQDVYRNFVTENTPHRKALVILGVHGNHLNDTLSYFLLNGTDLFHMIESAGVL 2021


gi|4885229|ref|
gi|14733833|ref
2050 2060 2070 2080 2090 2100
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2110 2120 2130 2140 2150 2160
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2170 2180 2190 2200 2210 2220
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2230 2240 2250 2260 2270 2280
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2290 2300 2310 2320 2330 2340
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2350 2360 2370 2380 2390 2400
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2410 2420 2430 2440 2450 2460
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2470 2480 2490 2500 2510 2520
NOV2
gi|13787217|ref
gi|7407144|gb|A
gi|12621132|ref
gi|4885229|ref|
gi|14733833|ref
2530 2540 2550 2560 2570 2580
NOV2




3070 3080 3090 3100 3110 3120
NOV2 DTDTNAQITYSLHGPGAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATDGGGRSCQAL 3098
gi|13787217|ref DTDTNAQITYSLHGPGAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATDGGGRSCQAL 3098
gi|7407144|gb|A DTDTNAQITYSLHGPGAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATDGGGRSCQAL 3098
gi|12621132|ref DMDTNAQITYSLHGPGAQEFKLDPHTGELTTLTVLDRERKDVFNLVAKATDGGGQSCQAE 3100
gi|4885229|ref| DIRSNAEITYQLLGSGAEKFKLNPDTGELKTSTPLDREQQAVYKLLVRATDGGGREGQAS 3112
gi|14733833|ref DTDTNAQITYSLHGPGAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATDGGGRSCQAL 1740
3130 3140 3150 3160 3170 3180
NOV2 ITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSAEG 3158
gi|13787217|ref ITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSAEG 3158
gi|7407144|gb|A ITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSAEG 3158
gi|12621132|ref VTLHVEDVNDNAPRFFPSHCDVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSADG 3160
gi|4885229|ref| IVVTLEDVNDNAPEFSADPYAITVPENTEPGTLLTRVQATDADASLNRKILYSLPDSADG 3172
gi|14733833|ref ITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQVVYSLPDSAEG 1800
3190 3200 3210 3220 3230 3240
NOV2 HFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPVF 3218
gi|13787217|ref HFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPVF 3218
gi|7407144|gb|A HFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPVF 3218
gi|12621132|ref QFSIDATSGVIRLEKPLQVRASSAVELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPIF 3220
gi|4885229|ref| QFSINELSGIIQLEKPLDRELQAVYTLSLKAVDQSLPRRLTATGTVIVSVLDINDNPPVF 3232
gi|14733833|ref HFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTPIPLSTLGTVTVSVVGLEDYLPVF 1860
3250 3260 3270 3280 3290 3300
NOV2 LNTEHSVQVPEDAPPGTEVLQLATLTR--PGAEKTGYRVVSGNEQGRFRLDARTGILYVN 3276
gi|13787217|ref LNTEHSVQVPEDAPPGTEVLQLATLTR--PGAEKTGYRVVSGNEQGRFRLDARTGILYVN 3276
gi|7407144|gb|A LNTEHSVQVPEDAPPGTEVLQLATLTR--PGAEKTGYRVVSGNEQGRFRLDARTGILYVN 3276
gi|12621132|ref LNAEHSTQVPEDAPIDMQVLQLATLTR--PGSEKTGYHLTGGNEQGKFRLDAHTGILYVN 3278
gi|4885229|ref| EYREYGATVSEDILVGTEVLQVYAASRDIEANAEITYSIISGNEHGKFSIDSKTGAVFII 3292
gi|14733833|ref LNTEHSVQVPEDAPPGTEVLQLATLTR--PGAEKTGYRVVSGNEQGRFRLDARTGILYVN 1918
3310 3320 3330 3340 3350 3360
NOV2 ASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNITDVNEHRPQFPQDPYSTRVLENALV 3336
gi|13787217|ref ASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNITDVNEHRPQFPQDPYSTRVLENALV 3336
gi|7407144|gb|A ASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNITDVNEHRPQFPQDPYSTRVLENALV 3336
gi|12621132|ref GSLDFETNPKYFLSIECSRKSSSSLSDVTTIVINVTDVNEHHPRFTHDLYTVRVLENAVV 3338
gi|4885229|ref| ENLDYESSHEYYLTVEATDGGTPSLSDVATVNVNVTDINDNTPVFSQDTYTIVVSEDAVI 3352
gi|14733833|ref ASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNITDVNEHRPQFPQDPYSTRVLENALV 1978
3370 3380 3390 3400 3410 3420
NOV2 GDVILTVSATDEDGPLNSDITYSLIGGNQLGHFTIHPKKGELQVAKALDREQASSYSLKL 3396
gi|13787217|ref GDVILTVSATDEDGPLNSDITYSLIGGNQLGHFTIHPKKGELQVAKALDREQASSYSLKL 3396
gi|7407144|gb|A GDVILTVSATDEDGPLNSDITYSLIGGNQLGHFTIHPKKGELQVAKALDREQASSYSLKL 3396
gi|12621132|ref GDVILTVSASDDDGPVNSAITYSLVGGNQLGHFTINPKKGKLQVAKALDWEQTPSYSLRL 3398
gi|4885229|ref| EQSVITVMADDADGPSNSHIHYSIIDGNQGSSFTIDPVRGEVKVVKKLLDREIISGYTLIV 3412
gi|14733833|ref GDVILTVSATDEDGPLNSDITYSLIGGNQLGHFTIHPKKGELQVAKALDREQASSYSLKL 2038
3430 3440 3450 3460 3470 3480
NOV2 RATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSP 3456
gi|13787217|ref RATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSP 3456
gi|7407144|gb|A RATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSP 3456
gi|12621132|ref RATDSGQPPLHEDTEVAVEVVDVNDNPPRFFQLNYSISVQENSPIGIKVLQLILEDPDSP 3458
gi|4885229|ref| QASDNGSPPRVNTTIVNIDVSDVNDNAPVFSRGNYSVIIQENKPVGFSVLQLVVIDEDSS 3472
gi|14733833|ref RATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSP 2098
3490 3500 3510 3520 3530 3540
NOV2 ENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTS 3516
gi|13787217|ref ENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTS 3516
gi|7407144|gb|A ENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTS 3516
gi|12621132|ref QNGPPYPFRITEGNTGSVFRVTPDGWLVTAASLSKKAREWYQLHIEVSDSGLPPLSSSTL 3518
gi|4885229|ref| HNGPPFPFTIVIGNDEKAFEVNPQGVLLISSAIKRKEKDHYLLQVKVADNGKPQLSSLTY 3532
gi|14733833|ref ENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTS 2158
3550 3560 3570 3580 3590 3600
NOV2 VRVHVTEQSHYAPSALPLEIFITVGEDEFQGGMVGKIHATDRDPQDTLTYSLAEEETLGF 3576
gi|13787217|ref VRVHVTEQSHYAPSALPLEIFITVGEDEFQGGMVGKIHATDRDPQDTLTYSLAEEETLGF 3576
gi|7407144|gb|A VRVHVTEQSHYAPSALPLEIFITVGEDEFQGGMVGKIHATDRDPQDTLTYSLAEEETLGR 3576
gi|12621132|ref VRVQVTEQSRYFPSTLPLEISLTKGSEEFQGGMIGKIHATDRDPQDTLTYSLEQEGGLDK 3578





Tables 2E list the domain description from DOMAIN analysis results against NOV2. This indicates that the NOV2 sequence has properties similar to those of other proteins known to contain this domain.

Table 2E. Domain Results for NOV2

PSSMs producing significant alignments:		Score (bits)	E value
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	97.8	1e-20
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	91.3	1e-18
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	89.7	3e-18
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	89.0	5e-18
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	89.0	5e-18
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	86.3	3e-17
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	84.3	1e-16
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	80.5	2e-15
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	75.9	5e-14
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	72.0	7e-13
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	72.0	7e-13
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	71.6	9e-13
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	71.6	9e-13
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	70.1	2e-12
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	69.7	3e-12
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	68.2	9e-12
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	66.6	3e-11
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	65.9	5e-11
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	65.1	8e-11
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	62.8	4e-10
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	61.2	1e-09
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	60.8	2e-09
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	60.1	3e-09
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	59.7	3e-09
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	55.5	6e-08
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	53.9	2e-07

gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	53.5	2e-07
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	53.1	3e-07
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	50.1	3e-06
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	46.2	4e-05
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	46.2	4e-05
gnl|Smart|smart00112	CA, Cadherin repeats.; Cadherins are glycoproteins involved in...	38.5	0.008
gnl|Pfam|pfam00028	cadherin, Cadherin domain	92.0	6e-19
gnl|Pfam|pfam00028	cadherin, Cadherin domain	85.9	4e-17
gnl|Pfam|pfam00028	cadherin, Cadherin domain	85.5	6e-17
gnl|Pfam|pfam00028	cadherin, Cadherin domain	80.5	2e-15
gnl|Pfam|pfam00028	cadherin, Cadherin domain	80.1	2e-15
gnl|Pfam|pfam00028	cadherin, Cadherin domain	79.7	3e-15
gnl|Pfam|pfam00028	cadherin, Cadherin domain	79.7	3e-15
gnl|Pfam|pfam00028	cadherin, Cadherin domain	79.7	3e-15
gnl|Pfam|pfam00028	cadherin, Cadherin domain	77.0	2e-14
gnl|Pfam|pfam00028	cadherin, Cadherin domain	76.3	3e-14
gnl|Pfam|pfam00028	cadherin, Cadherin domain	75.9	5e-14
gnl|Pfam|pfam00028	cadherin, Cadherin domain	74.7	1e-13
gnl|Pfam|pfam00028	cadherin, Cadherin domain	67.0	2e-11
gnl|Pfam|pfam00028	cadherin, Cadherin domain	66.6	3e-11
gnl|Pfam|pfam00028	cadherin, Cadherin domain	64.7	1e-10
gnl|Pfam|pfam00028	cadherin, Cadherin domain	64.3	1e-10
gnl|Pfam|pfam00028	cadherin, Cadherin domain	63.9	2e-10
gnl|Pfam|pfam00028	cadherin, Cadherin domain	59.3	4e-09

The above domains are located at amino acids 67-146, 170-254, 387-456, 463-541, 480-562, 587-661, 721-810, 737-818, 825-916, 842-923, 934-1021, 949-1026, 1038-1128, 1054-1135, 1145-1233, 1161-1240, 1247-1335, 1266-1335, 1374-1446, 1470-1553, 1577-1658, 1560-1650, 1688-1756, 1763-1862, 1787-1870, 1894-1963, 1998-2068, 2079-2163, 2092-2163, 2193-2270, 2277-2369, 2296-2377, 2401-2479, 2486-2576, 2505-2583, 2607-2681, 2716-2795, 2802-2897, 2819-2904, 2932-3009, 3016-3104, 3033-3111, 3120-3195,3135-3195, 3224-3312, 3253-3319, 3326-3416, 3343-3424, 3451-3529, and 3431-3522 of NOV2. Cadherins are glycoproteins involved in Ca2+-mediated cell-cell adhesion. Cadherin domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium.

Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor has homology to the b-catenin binding regions of classical cadherin cytoplasmic tails and also ends with a PDZ domain-binding motif {mu}-protocadherin that regulates branching morphogenesis in the kidneys and lungs. Therefore, NOV2 has a role in cell growth and cell survival. Therapeutic targeting of NOV2 with a monoclonal antibody is anticipated to limit or block the extent of cell growth and cell survival in colon, breast, liver and gastric tumors.

The disclosed NOV2 nucleic acid of the invention encoding a Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor-like protein includes the nucleic acid whose sequence is provided in Table 2A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 2A while still encoding a protein that maintains its Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV2 protein of the invention includes the Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor -like protein whose sequence is provided in Table 2B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 2B while still encoding a protein that maintains its Protocadherin Fat 2 (FAT2) cadherin related tumor suppressor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 56% percent of the residues may be so changed.

NOV2 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which are useful in understanding of pathology of the disease and development of new drug targets for various disorders. These antibodies could also be used to treat certain pathologies as detailed above.

NOV3

A disclosed NOV3 nucleic acid of 3381 nucleotides (also referred to as CG-SC 17661211) encoding a novel orphan GPCR-like protein is shown in Table 3A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 62-64 and ending with a TGA codon at nucleotides 2882-2884. The start and stop codons are in bold letters, and the 5' and 3' untranslated regions are underlined.

Table 3A. NOV3 Nucleotide Sequence (SEQ ID NO:5)

CTAGAATTCAGCGGCCGCTTAATTCAGAACGGCCCCCTGCCACTGCCAGGAGGACGGCATCATGCTGTCTG
CCGACTGCTCTGAGCTCGGGCTGTCCGCCGTTCCGGGGGACCCGGACCCCCTGACGGCTTACCTGGACCTC
AGCATGAACAACCTCACAGAGCTTCAGCCTGGCCTCTTCCACCACCTGCGCTTCTTGGAGGAGCTGCGTCT
CTCTGGGAACCATCTCTCACACATCCCAGGACAAGCATTCTCTGGTCTCTACAGCCTGAAAATCCTGATGC
TGCAGAACAATCAGCTGGGAGGAATCCCCGCAGAGGCGCTGTGGGAGCTGCCGAGCCTGCAGTCGCTGCGC
CTAGATGCCAACCTCATCTCCCTGGTCCCGGAGAGGAGCTTTGAGGGGCTGTCCTCCCTCCGCCACCTCTG
GCTGGACGACAATGCACTCACGGAGATCCCTGTCAGGGCCCTCAACAACCTCCCTGCCCTGCAGGCCATGA
CCCTGGCCCTCAACCGCATCAGCCACATCCCCGACTACGCGTTCCAGAATCTCACCAGCCTTGTGGTGCTG
CATTTGCATAACAACCGCATCCAGCATCTGGGGACCCACAGCTTCGAGGGGCTGCACAATCTGGAGACACT
AGACCTGAATTATAACAAGCTGCAGGAGTTCCCTGTGGCCATCCGGACCCTGGGCAGACTGCAGGAACTGG
GGTTCCATAACAACAACATCAAGGCCATCCCAGAAAAGGCCTTCATGGGGAACCCTCTGCTACAGACGATA
CACTTTTATGATAACCCAATCCAGTTTGTGGGAAGATCGGCATTCCAGTACCTGCCTAAACTCCACACACT
ATCTCTGAATGGTGCCATGGACATCCAGGAGTTTCCAGATCTCAAAGGCACCACCAGCCTGGAGATCCTGA
CCCTGACCCGCGCAGGCATCCGGCTGCTCCCATCGGGGATGTGCCAACAGCTGCCCAGGCTCCGAGTCCTG
GAACTGTCTCACAATCAAATTGAGGAGCTGCCCAGCCTGCACAGGTGTCAGAAATTGGAGGAAATCGGCCT
CCAACACAACCGCATCTGGGAAATTGGAGCTGACACCTTCAGCCAGCTGAGCTCCCTGCAAGCCCTGGATC
TTAGCTGGAACGCCATCCGGTCCATCCACCCCGAGGCCTTCTCCACCCTGCACTCCCTGGTCAAGCTGGAC
CTGACAGACAACCAGCTGACCACACTGCCCCTGGCTGGACTTGGGGGCTTGATGCATCTGAAGCTCAAAGG
GAACCTTGCTCTCTCCCAGGCCTTCTCCAAGGACAGTTTCCCAAAACTGAGGATCCTGGAGGTGCCTTATG
CCTACCAGTGCTGTCCCTATGGGATGTGTGCCAGCTTCTTCAAGGCCTCTGGGCAGTGGGAGGCTGAAGAC
CTTCACCTTGATGATGAGGAGTCTTCAAAAAGGCCCCTGGGCCTCCTTGCCAGACAAGCAGAGAACCACTA
TGACCAGGACCTGGATGAGCTCCAGCTGGAGATGGAGGACTCAAAGCCACACCCCAGTGTCCAGTGTAGCC
CTACTCCAGGCCCCTTCAAGCCCTGTGAGTACCTCTTTGAAAGCTGGGGCATCCGCCTGGCCGTGTGGGCC
ATCGTGTTGCTCTCCGTGCTCTGCAATGGACTGGTGCTGCTGACCGTGTTCGCTGGCGGGCCTGCCCCCCT
GCCCCCGGTCAAGTTTGTGGTAGGTGCGATTGCAGGCGCCAACACCTTGACTGGCATTTCCTGTGGCCTTC
TAGCCTCAGTCGATGCCCTGACCTTTGGTCAGTTCTCTGAGTACGGAGCCCGCTGGGAGACGGGGCTAGGC
TGCCGGGCCACTGGCTTCCTGGCAGTACTTGGGTCGGAGGCATCGGTGCTGCTGCTCACTCTGGCCGCAGT
GCAGTGCAGCGTCTCCGTCTCCTGTGTCCGGGCCTATGGGAAGTCCCCCTCCCTGGGCAGCGTTCGAGCAG
GGGTCCTAGGCTGCCTGGCACTGGCAGGGCTGGCCGCCGCACTGCCCCTGGCCTCAGTGGGAGAATACGGG
GCCTCCCCACTCTGCCTGCCCTACGCGCCACCTGAGGGTCAGCCAGCAGCCCTGGGCTTCACCGTGGCCCT
GGTGATGATGAACTCCTTCTGTTTCCTGGTCGTGGCCGGTGCCTACATCAAACTGTACTGTGACCTGCCGC
GGGGCGACTTTGAGGCCGTGTGGGACTGCGCCATGGTGAGGCACGTGGCCTGGCTCATCTTCGCAGACGGG
CTCCTCTACTGTCCCGTGGCCTTCCTCAGCTTCGCCTCCATGCTGGGCCTCTTCCCTGTCACGCCCGAGGC
CGTCAAGTCTGTCCTGCTGGTGGTGCTGCCCCTGCCTGCCTGCCTCAACCCACTGCTGTACCTGCTCTTCA
ACCCCCACTTCCGGGATGACCTTCGGCGGCTTCGGCCCCGCGCAGGGGACTCAGGGCCCCTAGCCTATGCT
GCGGCCGGGGAGCTGGAGAAGAGCTCCTGTGATTCTACCCAGGCCCTGGTAGCCTTCTCTGATGTGGATCT
CATTCTGGAAGCTTCTGAAGCTGGGCGGCCCCCTGGGCTGGAGACCTATGGCTTCCCCTCAGTGACCCTCA
TCTCCTGTCAGCAGCCAGGGGCCCCCAGGCTGGAGGGCAGCCATTGTGTAGAGCCAGAGGGGAACCACTTT
GGGAACCCCCAACCCTCCATGGATGGAGAACTGCTGCTGAGGGCAGAGGGATCTACGCCAGCAGGTGGAGG
CTTGTCAGGGGGTGGCGCTTTCAGCCCTCTGGCTTGGCCTTTGCTTCACACGTGTAAATATCCCTCCCCAT
TCTTCTCTTCCCCTCTCTTCCCTTTCCTCTCTCCCCCTCGGTGAATGATGGCTGCTTCTAAAACAAATACA
ACCAAAACTCAGCAGTGTGATCTATAGCAGGATGGCCCAGTACCTGGCTCCACTGATCACCTCTCTCCTGT
GACCATCACCAACGGGTGCCCTCTTGGCCTGGCTTTCCCTTGGCCTTCCTCAGCTTCACCTTGATACTGGG
CCTCTTCCTTGTCATGTCTGAAGCTGTGGACCAGAGACCTGGACTTTTGTCTGCTTAAGGGAAATGAGGGA
AGTAAAGACAGTGAAGGGGTGGAGGGTTGATCAGGGCACAGTGGACAGGGAGACCTCACAGAGAAAGGCCT
GGAAGGTGATTTCCCGTGTGACTCATGGATAGGATACAAAATGTGTTCCATGTACCATTAATCTTGACATA
TGCCATGCATAAAGACTTCCTATTAAAATAAGCTTTGGAAGAGATTACACATGATGTCTTTTTCTTAGAGA
TTCACAGTGCATGTTAGTGTAATAAAGAGATAAGTCCTACAGTA

The disclosed NOV3 nucleic acid sequence has 1657 of 1659 bases (99%) identical to the 3119 nucleotide *Homo sapiens* VTS20631 mRNA, g-protein coupled receptor family partial cds (GENBANK-ID: gi|13447609|dbj|AB049405.1|AB049405) (E = 0.0).

A disclosed NOV3 protein (SEQ ID NO:6) encoded by SEQ ID NO:5 has 940 amino acid residues, and is presented using the one-letter code in Table 3B. Signal P, Psort and/or Hydropathy results predict that NOV3 does not have a signal peptide, and is likely to be localized to the plasma membrane as a Type IIIb membrane protein.

Table 3B. Encoded NOV3 protein sequence (SEQ ID NO:6).

```
MLSADCSELGLSAVPGDPDPLTAYLDLSMNNLTELQPGLFHHLRFLEELRLSGNHLSHIPGQAFSGLYSLK
ILMLQNNQLGGIPAEALWELPSLQSLRLDANLISLVPERSFEGLSSLRHLWLDDNALTEIPVRALNNLPAL
QAMTLALNRISHIPDYAFQNLTSLVVLHLHNNRIQHLGTHSFEGLHNLETLDLNYNKLQEFPVAIRTLGRL
QELGFHNNNIKAIPEKAFMGNPLLQTIHFYDNPIQFVGRSAFQYLPKLHTLSLNGAMDIQEFPDLKGTTSL
EILTLTRAGIRLLPSGMCQQLPRLRVLELSHNQIEELPSLHRCQKLEEIGLQHNRIWEIGADTFSQLSSLQ
ALDLSWNAIRSIHPEAFSTLHSLVKLDLTDNQLTTLPLAGLGGLMHLKLKGNLALSQAFSKDSFPKLRILE
VPYAYQCCPYGMCASFFKASGQWEAEDLHLDDEESSKRPLGLLARQAENHYDQDLDELQLEMEDSKPHPSV
QCSPTPGPFKPCEYLFESWGIRLAVWAIVLLSVLCNGLVLLTVFAGGPAPLPPVKFVVGAIAGANTLTGIS
CGLLASVDALTFGQFSEYGARWETGLGCRATGFLAVLGSEASVLLLTLAAVQCSVSVSCVRAYGKSPSLGS
VRAGVLGCLALAGLAAALPLASVGEYGASPLCLPYAPPEGQPAALGFTVALVMMNSFCFLVVAGAYIKLYC
DLPRGDFEAVWDCAMVRHVAWLIFADGLLYCPVAFLSFASMLGLFPVTPEAVKSVLLVVLPLPACLNPLLY
LLFNPHFRDDLRRLRPRAGDSGPLAYAAAGELEKSSCDSTQALVAFSDVDLILEASEAGRPPGLETYGFPS
VTLISCQQPGAPRLEGSHCVEPEGNHFGNPQPSMDGELLLRAEGSTPAGGGLSGGGAFSPLAWPLLHTCKY
PSPFFSSPLFPFLSPPR
```

TaqMan expression data for NOV3 is found below is Example 1, and SAGE data is found below in Example 2. The TaqMan data indicates overexpression of NOV3 in colon, breast, liver and gastric tumors.

NOV3 has homology to the amino acid sequences shown in the BLASTP data listed in Table 3C.

Table 3C. BLAST results for NOV3

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|13447610|dbj|BAB 39854.1| (AB049405)	VTS20631 [Homo sapiens]	928	802/895 (89%)	802/895 (89%)	0.0
gi|15298008|ref|XP_ 046692.2|	similar to leucine-rich repeat-containing G protein-coupled receptor 6 (H. sapiens) [Homo sapiens]	893	774/867 (89%)	774/867 (89%)	0.0

gi|10441732|gb|AAG17168.1|AF190501_1 (AF190501)	leucine-rich repeat-containing G protein-coupled receptor 6 [Homo sapiens]	828	638/798 (79%)	653/798 (80%)	0.0
gi|4504379|ref|NP_003658.1|	G protein-coupled receptor 49; orphan G protein-coupled receptor HG38; G protein-coupled receptor 67 [Homo sapiens]	907	436/869 (50%)	556/869 (63%)	0.0
gi|3885472|gb|AAC77911.1| (AF061444)	G protein-coupled receptor LGR5 [Homo sapiens]	907	434/869 (49%)	554/869 (62%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 3D.

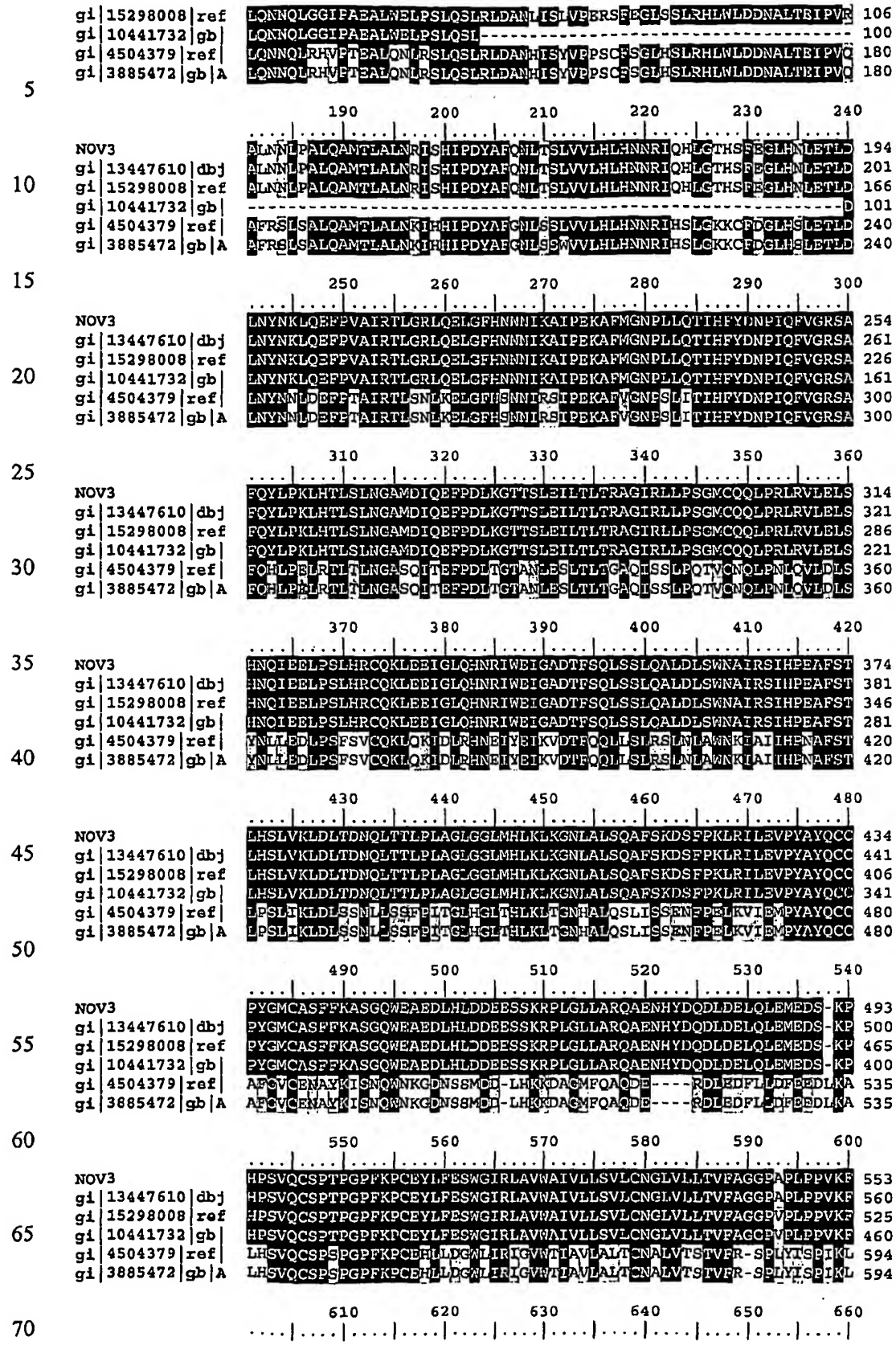
Table 3D. ClustalW Analysis of NOV3

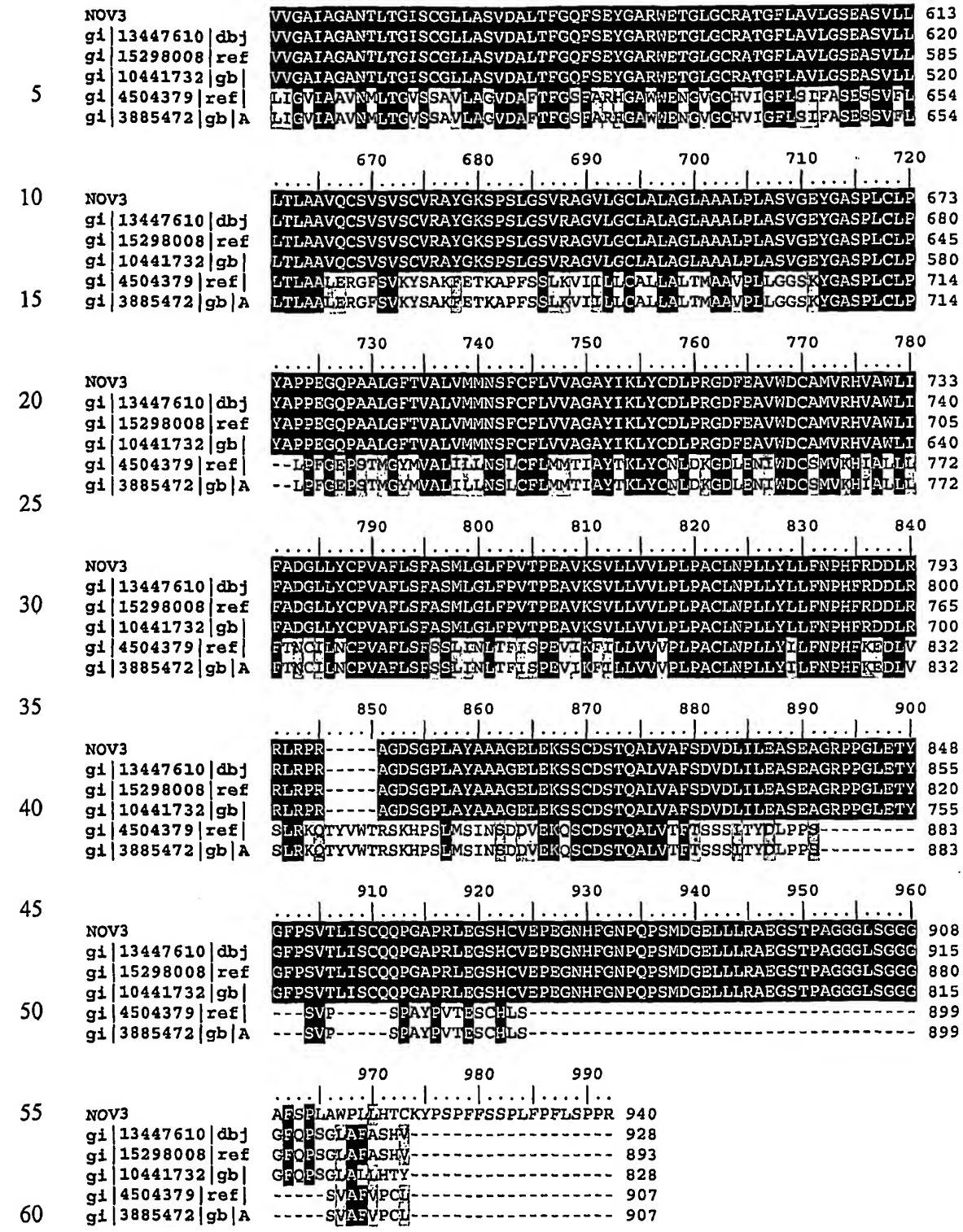
1) NOV3 (SEQ ID NO:6)
2) gi|13447610|dbj|BAB39854.1| (AB049405) VTS20631 [Homo sapiens] (SEQ ID NO:23)
3) gi|15298008|ref|XP_046692.2| similar to leucine-rich repeat-containing G protein-coupled receptor 6 (H. sapiens) [Homo sapiens] (SEQ ID NO:24)
4) gi|10441732|gb|AAG17168.1|AF190501_1 (AF190501) leucine-rich repeat-containing G protein-coupled receptor 6 [Homo sapiens] (SEQ ID NO:25)
5) gi|4504379|ref|NP_003658.1| G protein-coupled receptor 49; orphan G protein-coupled receptor HG38; G protein-coupled receptor 67 [Homo sapiens] (SEQ ID NO:26)
6) gi|3885472|gb|AAC77911.1| (AF061444) G protein-coupled receptor LGR5 [Homo sapiens] (SEQ ID NO:27)

```
                        10        20        30        40        50        60
                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV3            ------------------------------------------MLSADCSLGLSAV 14
gi|13447610|dbj ----------------------------------HCQEDGIMLSADCSLGLSAV 21
gi|15298008|ref ------------------------------------------------------- 1
gi|10441732|gb| ------------------------------------------MRLEGEGRSARAC 17
gi|4504379|ref| MDTSRLGVLLSLPVLLQLATGGSSPRSGVLLRGCPTHCHCEPDGRMLLRVDCSDLGLSEL 60
gi|3885472|gb|A MDTSRLGVLLSLPVLLQLATGGSSPRSGVLLRGCPTHCHCEPDGRMLLRVDCSDLGLSEL 60

                        70        80        90       100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV3            PGDPDPLTAYLDLSMNNLTELQPGLFHHLRFLEELRLSGNHLSHIPGQAFSGLYSLKILM 74
gi|13447610|dbj PGDLDPLTAYLDLSMNNLTELQPGLFHHLRFLEELRLSGNHLSHIPGQAFSGLYSLKILM 81
gi|15298008|ref ------------MNNLTELQPGLFHHLRFLEELRLSGNHLSHIPGQAFSGLYSLKILM 46
gi|10441732|gb| RAGSARRGAPRDLSMNNLTELQPGLFHHLRFLEELRLSGNHLSHIPGQAFSGLYSLKILM 77
gi|4504379|ref| PSNLSVFTSYLDLSMNNISQLLPNPLPSLRFLEELRLAGNALTYIPKGAFTGLYSLKVLM 120
gi|3885472|gb|A PSNLSVFTSYLDLSMNNISQLLPNPLPSLHFLEELRLAGNALTYIPKGAFTGLYSLKVLM 120

                       130       140       150       160       170       180
                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV3            LQNNQLGGIPAEALWELPSLQSLRLDANLISLVPERSFEGLSSLRHLWLDDNALTEIPVR 134
gi|13447610|dbj LQNNQLGGIPAEALWELPSLQSLRLDANLISLVPERSFEGLSSLRHLWLDDNALTEIPVR 141
```


gi|15298008|ref
gi|10441732|gb|
gi|4504379|ref|
gi|3885472|gb|A
NOV3
gi|13447610|dbj
gi|15298008|ref
gi|10441732|gb|
gi|4504379|ref|
gi|3885472|gb|A



According to InterPro Domains searches, NOV3 contains 16 Leucine Rich Repeats domains and 2 seven transmembrane receptor (rhodopsin) domains.

Because of its high homology to GPCRs and its containing GPCR 7 transmembrane domains, NOV3 is thought to be involved with cell growth and cell survival. Therapeutic

targeting of NOV3 with a monoclonal antibody is anticipated to limit or block the extent of cell growth and cell survival in colon, breast, liver and gastric tumors.

The disclosed NOV3 nucleic acid of the invention encoding a Orphan GPCR-like protein includes the nucleic acid whose sequence is provided in Table 3A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 3A while still encoding a protein that maintains its Orphan GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV3 protein of the invention includes the Orphan GPCR -like protein whose sequence is provided in Table 3B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 3B while still encoding a protein that maintains its Orphan GPCR -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 51% percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the Orphan GPCR-like protein and nucleic acid (NOV3) disclosed herein suggest that NOV3 may have important structural and/or physiological functions characteristic of the citron kinase-like family. Therefore, the NOV3 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

NOV3 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. This novel protein also has value in development of powerful assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.These antibodies could also be used to treat certain pathologies as detailed above.

NOV4

A disclosed NOV4 nucleic acid of 2397 nucleotides (designated CuraGen Acc. No. CG-SC28471525) encoding a novel Slit-like protein is shown in Table 4A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TAG codon at nucleotides 2395-2397. In Table 4A the start and stop codons are in bold letters.

Table 4A. NOV4 Nucleotide Sequence (SEQ ID NO:7)

```
ATGCTAATAAATTGTGAAGCAAAAGGTATCAAGATGGTATCTGAAATAAGTGTGCCACCATCACGACCTT
TCCAACTAAGCTTATTAAATAACGGCTTGACGATGCTTCACACAAATGACTTTTCTGGGCTTACCAATGC
TATTTCAATACACCTTGGATTTAACAATATTGCAGATATTGAGATAGGTGCATTTAATGGCCTTGGCCTC
CTGAAACAACTTCATATCAATCACAATTCTTTAGAAATTCTTAAAGAGGATACTTTCCATGGACTGGAAA
ACCTGGAATTCCTGCAAGCAGATAACAATTTTATCACAGTGATTGAACCAAGTGCCTTTAGCAAGCTCAA
CAGACTCAAAGTGTTAATTTTAAATGACAATGCTATTGAGAGTCTTCCTCCAAACATCTTCCGATTTGTT
CCTTTAACCCATCTAGATCTTCGTGGAAATCAATTACAAACATTGCCTTATGTTGGTTTTCTCGAACACA
TTGGCCGAATATTGGATCTTCAGTTGGAGGACAACAAATGGGCCTGCAATTGTGACTTATTGCAGTTAAA
AACTTGGTTGGAGAACATGCCTCCACAGTCTATAATTGGTGATGTTGTCTGCAACAGCCCTCCATTTTTT
AAAGGAAGTATACTCAGTAGACTAAAGAAGGAATCTATTTGCCCTACTCCACCAGTGTATGAAGAACATG
AGGATCCTTCAGGATCATTACATCTGGCAGCAACATCTTCAATAAATGATAGTCGCATGTCAACTAAGAC
CACGTCCATTCTAAAACTACCCACCAAAGCACCAGGTTTGATACCTTATATTACAAAGCCATCCACTCAA
CTTCCAGGACCTTACTGCCCTATTCCTTGTAACTGCAAAGTCCTATCCCCATCAGGACTTCTAATACATT
GTCAGGAGCGCAACATTGAAAGCTTATCAGATCTGAGACCTCCTCCGCAAAATCCTAGAAAGCTCATTCT
AGCGGGAAATATTATTCACAGTTTAATGAAGTCTGATCTAGTGGAATATTTCACTTTGGAAATGCTTCAC
TTGGGAAACAATCGTATTGAAGTTCTTGAAGAAGGATCGTTTATGAACCTAACGAGATTACAAAAACTCT
ATCTAAATGGTAACCACCTGACCAAATTAAGTAAAGGCATGTTCCTTGGTCTCCATAATCTTGAATACTT
ATATCTTGAATACAATGCCATTAAGGAAATACTGCCAGGAACCTTTAATCCAATGCCTAAACTTAAAGTC
CTGTATTTAAATAACAACCTCCTCCAAGTTTTACCACCACATATTTTTTCAGGGGTTCCTCTAACTAAGG
TAAATCTTAAAACAAACCAGTTTACCCATCTACCTGTAAGTAATATTTTGGATGATCTTGATTTGCTAAC
CCAGATTGACCTTGAGGATAACCCCTGGGACTGCTCCTGTGACCTGGTTGGACTGCAGCAATGGATACAA
AAGTTAAGCAAGAACACAGTGACAGATGACATCCTCTGCACTTCCCCCGGGCATCTCGACAAAAAGGAAT
TGAAAGCCCTAAATAGTGAAATTCTCTGTCCAGGTTTAGTAAATAACCCATCCATGCCAACACAGACTAG
TTACCTTATGGTCACCACTCCTGCAACAACAACAAATACGGCTGATACTATTTTACGATCTCTTACGGAC
GCTGTGCCACTGTCTGTTCTAATATTGGGACTTCTGATTATGTTCATCACTATTGTTTTCTGTGCTGCAG
GGATAGTGGTTCTTGTTCTTCACCGCAGGAGAAGATACAAAAAGAAACAAGTAGATGAGCAAATGAGAGA
CAACAGTCCTGTGCATCTTCAGTACAGCATGTATGGCCATAAAACCACTCATCACACTACTGAAAGACCC
TCTGCCTCACTCTATGAACAGCACATGGTGAGCCCCATGGTTCATGTCTATAGAAGTCCATCCTTTGGTC
CAAAGCATCTGGAAGAGGAAGAAGAGAGGAATGAGAAAGAAGGAAGTGATGCAAAACATCTCCAAAGAAG
TCTTTTGGAACAGGAAAATCATTCACCACTCACAGGGTCAAATATGAAATACAAAACCACGAACCAATCA
ACAGAATTTTTATCCTTCCAAGATGCCAGCTCATTGTACAGAAACATTTTAGAAAAAGAAAGGGAACTTC
AGCAACTGGGAATCACAGAATACCTAAGGAAAAACATTGCTCAGCTCCAGCCTGATATGGAGGCACATTA
TCCTGGAGCCCACGAAGAGCTGAAGTTAATGGAAACATTAATGTACTCACGTCCAAGGAAGGTATTAGTG
GAACAGACAAAAAATGAGTATTTTGAACTTAAAGCTAATTTACATGCTGAACCTGACTATTTAGAAGTCC
TGGAGCAGCAAACATAG
```

The nucleic acid sequence of NOV4, located on chromosome 13, has 2397 of 2397 bases (100%) identical to a 2593 nucleotide *Homo sapiens* hypothetical protein FLJ22774 (FLJ22774), mRNA (GENBANK-ID: gi|14758125|ref|XM_033182.1|) (E = 0.0).

A NOV4 polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 is 798 amino acid residues and is presented using the one letter code in Table 4B. Signal P, Psort and/or Hydropathy results predict that NOV4 is likely to be localized at the plasma membrane and is a Type Ib transmembrane protein.

Table 4B. NOV4 protein sequence (SEQ ID NO:8)

```
MLINCEAKGIKMVSEISVPPSRPFQLSLLNNGLIMLHTNDFSGLTNAISIHLGFNNIADIEIGAFNGLGLLKQL
HINHNSLEILKEDTFHGLENLEFLQADNNFITVIEPSAFSKLNRLKVLILNDNAIESLPPNIFRFVPLTHLDLR
GNQLQTLPYVGFLEHIGRILDLQLEDNKWACNCDLLQLKTWLENMPPQSIIGDVVCNSPPFFKGSILSRLKKES
ICPTPPVYEEHEDPSGSLHLAATSSINDSRMSTKTTSILKLPTKAPGLIPYITKPSTQLPGPYCPIPCNCKVLS
PSGLLIHCQERNIESLSDLRPPPQNPRKLILAGNIIHSLMKSDLVEYFTLEMLHLGNNRIEVLEEGSFMNLTRL
QKLYLNGNHLTKLSKGMFLGLHNLEYLYLEYNAIKEILPGTFNPMPKLKVLYLNNNLLQVLPPHIFSGVPLTKV
NLKTNQFTHLPVSNILDDLDLLTQIDLEDNPWDCSCDLVGLQQWIQKLSKNTVTDDILCTSPGHLDKKELKALN
SEILCPGLVNNPSMPTQTSYLMVTTPATTTNTADTILRSLTDAVPLSVLILGLLIMFITIVFCAAGIVVLVLHR
RRRYKKKQVDEQMRDNSPVHLQYSMYGHKTTHHTTERPSASLYEQHMVSPMVHVYRSPSFGPKHLEEEEERNEK
EGSDAKHLQRSLLEQENHSPLTGSNMKYKTTNQSTEFLSFQDASSLYRNILEKERELQQLGITEYLRKNIAQLQ
```

PDMEAHYPGAHEELKLMETLMYSRPRKVLVEQTKNEYPELKANLHAEPDYLEVLEQQT*

The full amino acid sequence of the protein of the invention was found to have 1263 of 1857 amino acid residues (68%) identical to, and 1501 of 1857 amino acid residues (80%) similar to, the 1884 amino acid residue Slit-2 protein from mouse (SPTREMBL-P70207) (E =0.0), and 364 of 801 amino acid residues (45%) identical to, and 520 of 801 amino acid residues (64%) similar to, the 2135 amino acid residue Human Slit protein (patp:AAU00019) ($E = 2.6^{-283}$).

The disclosed NOV4 protein is expressed in at least the following tissues: fibroblast like synoviocytes (normal), fetal brain, adipose, microvascular endothelial cells-lung, thalamus, fetal cerebral cortex, temporal lobe, parietal lobe, fetal cerebellum, and testis. TaqMan expression data for NOV4 is shown below in Example 1 and SAGE data is shown below in Example 2. The TaqMan data shows overexpression in several cell lines, especially those derived from brain tumors, metastatic breast and bladder tumors. EST analysis showed expression of NOV2 in neuroendocrine lung carcinoid and Endometrial tumor, plus 2 annotated as breast and bladder tumors.

NOV4 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 4C.

Table 4C. BLAST results for NOV4

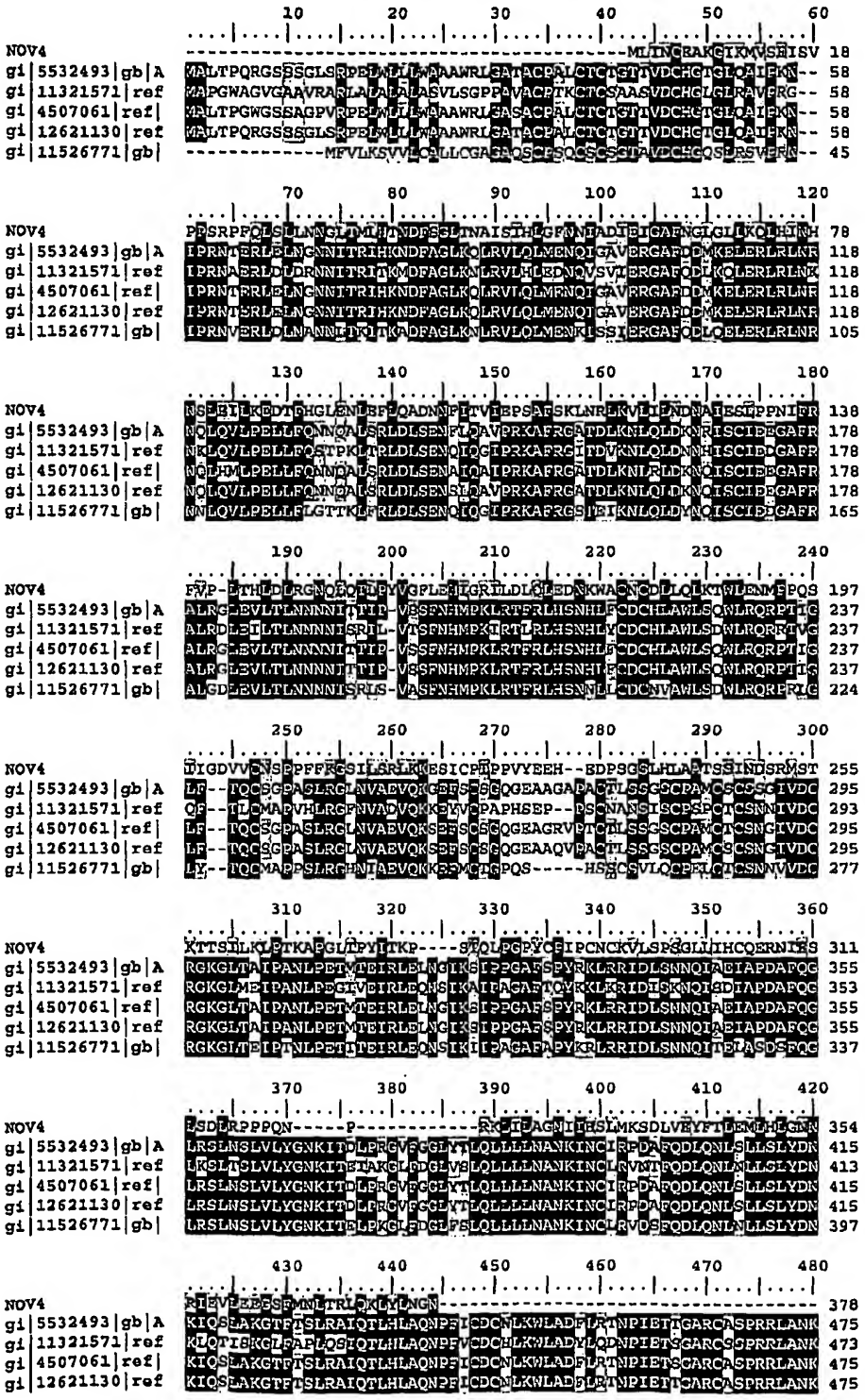
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|5532493|gb|AAD44758.1|AF144627_1 (AF144627)	SLIT1 [Mus musculus]	1531	123/520 (23%)	194/520 (36%)	5e-25
gi|11321571|ref|NP_003053.1|	slit (Drosophila) homolog 3; slit2; slit (Drosophila) homolog 2 [Homo sapiens]	1523	128/525 (24%)	202/525 (38%)	7e-25
gi|4507061|ref|NP_003052.1|	slit (Drosophila) homolog 1; slit1 [Homo sapiens]	1534	120/519 (23%)	190/519 (36%)	7e-24
gi|12621130|ref|NP_075242.1|	Slit1 [Rattus norvegicus]	1531	120/519 (23%)	191/519 (36%),	8e-24
gi|11526771|gb|AAG36773.1| (AF210321)	Slit2 [Danio rerio]	1512	132/531 (24%)	199/531 (36%)	1e-23

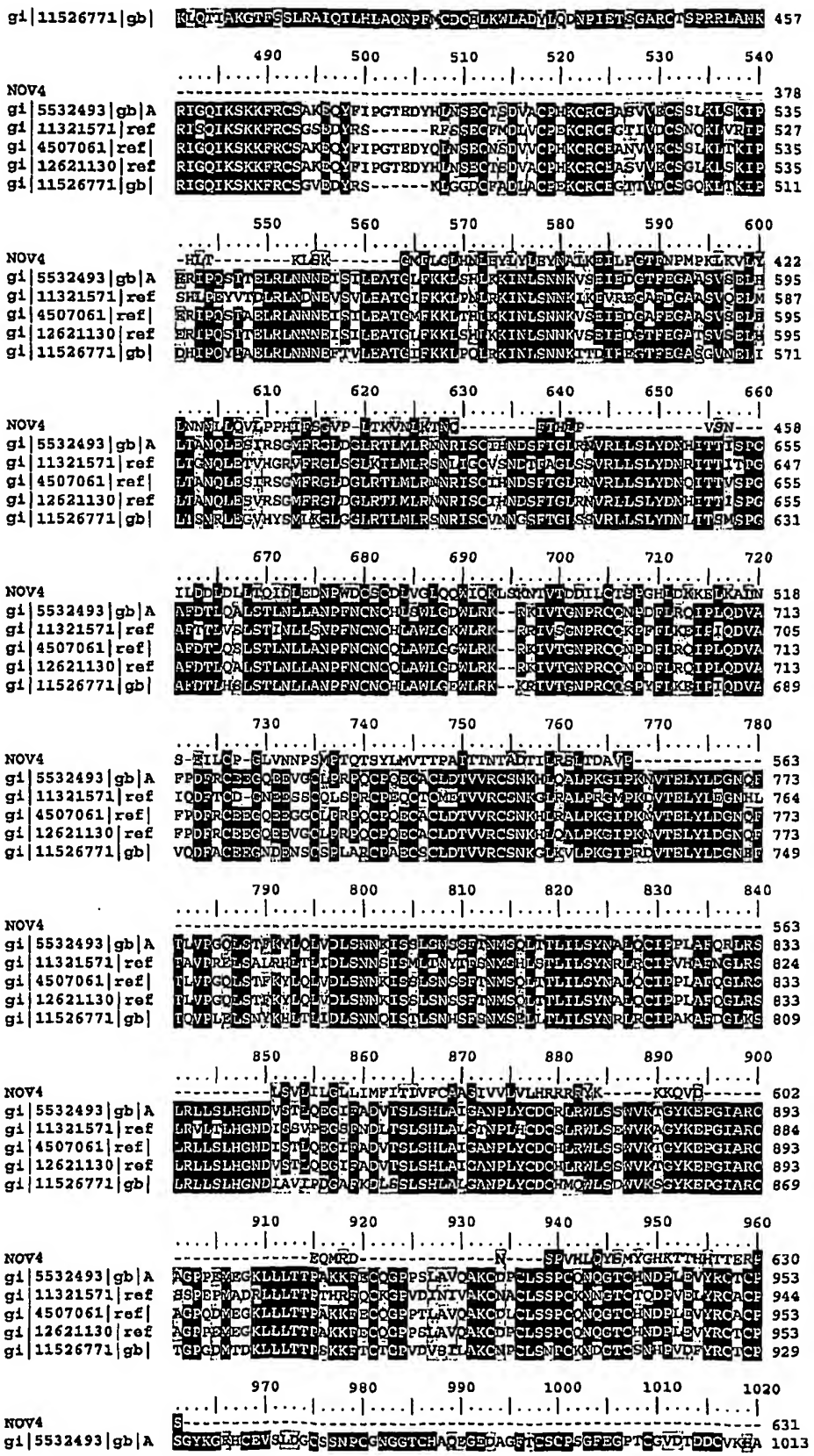
The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 4D.

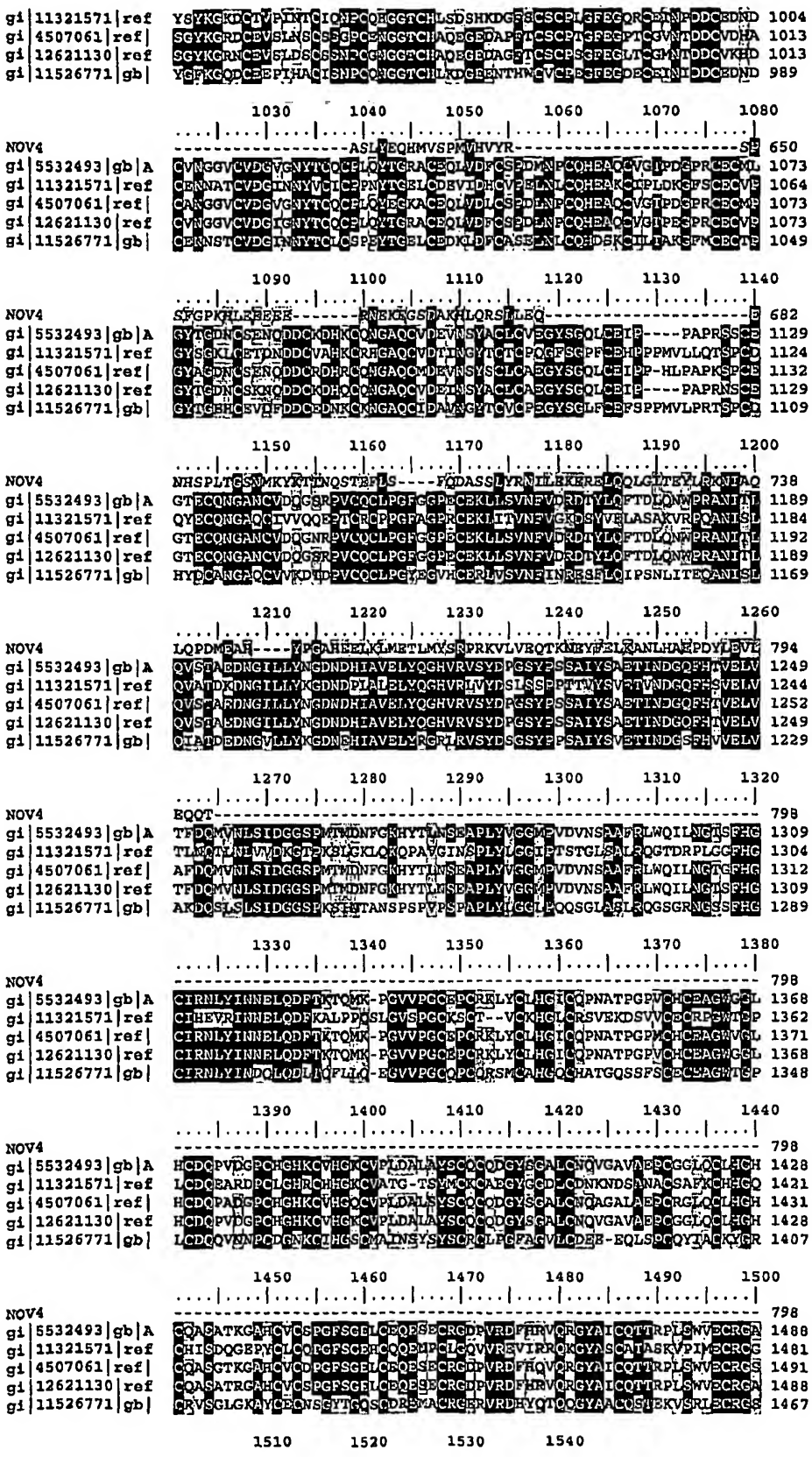
Table 4D ClustalW Analysis of NOV4

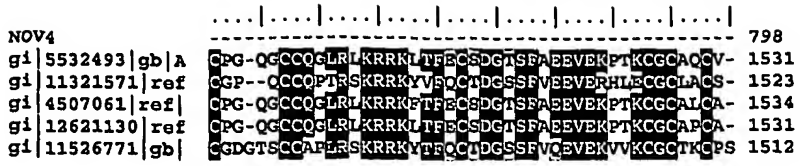
1) NOV4 (SEQ ID NO:8)
2) gi|5532493|gb|AAD44758.1|AF144627_1 (AF144627) SLIT1 [Mus musculus] (SEQ ID NO:28)

3) gi|11321571|ref|NP_003053.1| slit (Drosophila) homolog 3; slit2; slit (Drosophila) homolog 2 [Homo sapiens] (SEQ ID NO:29)
4) gi|4507061|ref|NP_003052.1| slit (Drosophila) homolog 1; slit1 [Homo sapiens] (SEQ ID NO:30)
5) gi|12621130|ref|NP_075242.1| Slit1 [Rattus norvegicus] (SEQ ID NO:31)
6) gi|11526771|gb|AAG36773.1| (AF210321) Slit2 [Danio rerio] (SEQ ID NO:32)









Tables 4E-H list the domain description from DOMAIN analysis results against NOV4. This indicates that the NOV4 sequence has properties similar to those of other proteins known to contain this domain.

Table 4E. Domain Analysis of NOV4

gnl|Smart|smart00082, LRRCT, Leucine rich repeat C-terminal domain
(SEQ ID NO:43)
CD-Length = 51 residues, 100.0% aligned
Score = 49.7 bits (117), Expect = 6e-07

```
Query:  474  NPWDCSCDLVGLQQWIQKLSKNTVTDDILCTSPGHLDKKELKALNSEILCP  524
             ||+ | |+|  | +| |          |+ | ||  |    |  | |   ||
Sbjct:  1    NPFICDCELRWLLRWLQANRHLQDPVDLRCASPESLRGPLLLLLPSSFKCP  51
```

Table 4F. Domain Analysis of NOV4

gnl|Smart|smart00082, LRRCT, Leucine rich repeat C-terminal domain
(SEQ ID NO:43)
CD-Length = 51 residues, 100.0% aligned
Score = 45.1 bits (105), Expect = 2e-05

```
Query:  175  NKWACNCDLLQLKTWLENMPPQSIIGDVVCNSPPFFKGSILSRLKKESICP  225
             | + |+|+|  |  ||+         |+ | ||   +| +|  |     ||
Sbjct:  1    NPFICDCELRWLLRWLQANRHLQDPVDLRCASPESLRGPLLLLLPSSFKCP  51
```

Table 4G. Domain Analysis of NOV4

gnl|Pfam|pfam01463, LRRCT, Leucine rich repeat C-terminal domain.
Leucine Rich Repeats pfam00560 are short sequence motifs present in a number of proteins with diverse functions and cellular locations. Leucine Rich Repeats are often flanked by cysteine rich domains. This domain is often found at the C-terminus of tandem leucine rich repeats. (SEQ ID NO:49)
CD-Length = 51 residues, 100.0% aligned
Score = 48.1 bits (113), Expect = 2e-06

```
Query:  474  NPWDCSCDLVGLQQWIQKLSKNTVTDDILCTSPGHLDKKELKALNSEILCP  524
             ||+ | |+|  | +|+++  +    +|+ | ||  |    |+ | |+  ||
Sbjct:  1    NPFICDCELRWLLRWLREPRRLEDPEDLRCASPESLRGPLLELLPSDFSCP  51
```

Table 4H. Domain Analysis of NOV4

gnl|Pfam|pfam01463, LRRCT, Leucine rich repeat C-terminal domain. Leucine Rich Repeats pfam00560 are short sequence motifs present in a number of proteins with diverse functions and cellular locations. Leucine Rich Repeats are often flanked by cysteine rich domains. This domain is often found at the C-terminus of tandem leucine rich repeats (SEQ ID NO:49)
CD-Length = 51 residues, 100.0% aligned
Score = 46.6 bits (109), Expect = 5e-06

```
Query:  175  NKWACNCDLLQLKTWLENMPPQSIIGDVVCNSPPFFKGSILSRLKKESICP  225
             | + |+|+|  |  ||          |+ | ||   +|  |  |  +   |
Sbjct:  1    NPFICDCELRWLLRWLREPRRLEDPEDLRCASPESLRGPLLELLPSDFSCP  51
```

NOV4 blocks Natriuretic peptide receptor proteins, possibly a receptor with ATP binding and Kinase activity. NOV4 is thought to be involved with metastatic potential. Therapeutic targeting of NOV4 with a monoclonal antibody is anticipated to limit or block the extent of metastasis in breast and brain tumors.

The disclosed NOV4 nucleic acid of the invention encoding a Slit-like protein includes the nucleic acid whose sequence is provided in Table 4A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 4A while still encoding a protein that maintains its Slit-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV4 protein of the invention includes the Slit-like protein whose sequence is provided in Table 4B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 4B while still encoding a protein that maintains its Slit -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 76% percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the Slit-like protein and nucleic acid (NOV4) disclosed herein suggest that this NOV4 protein may have important structural and/or physiological functions characteristic of the Slit family. Therefore, the NOV4 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

NOV4 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders. These antibodies could also be used to treat certain pathologies as decribed above.

NOV5

A disclosed NOV5 nucleic acid of 3825 nucleotides (also referred to as AC133) encoding a novel AC133 antigen-like protein is shown in Table 5A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 69-71 and ending with a TGA codon at nucleotides 2664-2666. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. NOV5 Nucleotide Sequence (SEQ ID NO:9)

GNNNNNNANNNNATTCNTNCANTGNACNNNACCAAGTTCTACCTCATGTTTGGAGG
ATCTTGCTAGCT**ATG**GCCCTCGTACTCGGCTCCCTGTTGCTGCTGGGGCTGTGCGG
GAACTCCTTTTCAGGAGGGCAGCCTTCATCCACAGATGCTCCTAAGGCTTGGAATT
ATGAATTGCCTGCAACAAATTATGAGACCCAAGACTCCCATAAAGCTGGACCCATT
GGCATTCTCTTTGAACTAGTGCATATCTTTCTCTATGTGGTACAGCCGCGTGATTT
CCCAGAAGATACTTTGAGAAAATTCTTACAGAAGGCATATGAATCCAAAATTGATT
ATGACAAGCCAGAAACTGTAATCTTAGGTCTAAAGATTGTCTACTATGAAGCAGGG
ATTATTCTATGCTGTGTCCTGGGGCTGCTGTTTATTATTCTGATGCCTCTGGTGGG
GTATTTCTTTTGTATGTGTCGTTGCTGTAACAAATGTGGTGGAGAAATGCACCAGC

GACAGAAGGAAAATGGGCCCTTCCTGAGGAAATGCTTTGCAATCTCCCTGTTGGTG
ATTTGTATAATAATAAGCATTGGCATCTTCTATGGTTTTGTGGCAAATCACCAGGT
AAGAACCCGGATCAAAAGGAGTCGGAAACTGGCAGATAGCAATTTCAAGGACTTGC
GAACTCTCTTGAATGAAACTCCAGAGCAAATCAAATATATATTGGCCCAGTACAAC
ACTACCAAGGACAAGGCGTTCACAGATCTGAACAGTATCAATTCAGTGCTAGGAGG
CGGAATTCTTGACCGACTGAGACCCAACATCATCCCTGTTCTTGATGAGATTAAGT
CCATGGCAACAGCGATCAAGGAGACCAAAGAGGCGTTGGAGAACATGAACAGCACC
TTGAAGAGCTTGCACCAACAAAGTACACAGCTTAGCAGCAGTCTGACCAGCGTGAA
AACTAGCCTGCGGTCATCTCTCAATGACCCTCTGTGCTTGGTGCATCCATCAAGTG
AAACCTGCAACAGCATCAGATTGTCTCTAAGCCAGCTGAATAGCAACCCTGAACTG
AGGCAGCTTCCACCCGTGGATGCAGAACTTGACAACGTTAATAACGTTCTTAGGAC
AGATTTGGATGGCCTGGTCCAACAGGGCTATCAATCCCTTAATGATATACCTGACA
GAGTACAACGCCAAACCACGACTGTCGTAGCAGGTATCAAAAGGGTCTTGAATTCC
ATTGGTTCAGATATCGACAATGTAACTCAGCGTCTTCCTATTCAGGATATACTCTC
AGCATTCTCTGTTTATGTTAATAACACTGAAAGTTACATCCACAGAAATTTACCTA
CATTGGAAGAGTATGATTCATACTGGTGGCTGGGTGGCCTGGTCATCTGCTCTCTG
CTGACCCTCATCGTGATTTTTTACTACCTGGGCTTACTGTGTGGCGTGTGCGGCTA
TGACAGGCATGCCACCCCGACCACCCGAGGCTGTGTCTCCAACACCGGAGGCGTCT
TCCTCATGGTTGGAGTTGGATTAAGTTTCCTCTTTTGCTGGATATTGATGATCATT
GTGGTTCTTACCTTTGTCTTTGGTGCAAATGTGGAAAAACTGATCTGTGAACCTTA
CACGAGCAAGGAATTATTCCGGGTTTTGGATACACCCTACTTACTAAATGAAGACT
GGGAATACTATCTCTCTGGGAAGCTATTTAATAAATCAAAAATGAAGCTCACTTTT
GAACAAGTTTACAGTGACTGCAAAAAAAATAGAGGCACTTACGGCACTCTTCACCT
GCAGAACAGCTTCAATATCAGTGAACATCTCAACATTAATGAGCATACTGGAAGCA
TAAGCAGTGAATTGGAAAGTCTGAAGGTAAATCTTAATATCTTTCTGTTGGGTGCA
GCAGGAAGAAAAAACCTTCAGGATTTTGCTGCTTGTGGAATAGACAGAATGAATTA
TGACAGCTACTTGGCTCAGACTGGTAAATCCCCCGCAGGAGTGAATCTTTTATCAT
TTGCATATGATCTAGAAGCAAAAGCAAACAGTTTGCCCCCAGGAAATTTGAGGAAC
TCCCTGAAAAGAGATGCACAAACTATTAAAACAATTCACCAGCAACGAGTCCTTCC
TATAGAACAATCACTGAGCACTCTATACCAAAGCGTCAAGATACTTCAACGCACAG
GGAATGGATTGTTGGAGAGAGTAACTAGGATTCTAGCTTCTCTGGATTTTGCTCAG
AACTTCATCACAAACAATACTTCCTCTGTTATTATTGAGGAAACTAAGAAGTATGG
GAGAACAATAATAGGATATTTTGAACATTATCTGCAGTGGATCGAGTTCTCTATCA
GTGAGAAAGTGGCATCGTGCAAACCTGTGGCCACCGCTCTAGATACTGCTGTTGAT
GTCTTTCTGTGTAGCTACATTATCGACCCCTTGAATTTGTTTTGGTTTGGCATAGG
AAAAGCTACTGTATTTTTACTTCCGGCTCTAATTTTTGCGGTAAAACTGGCTAAGT
ACTATCGTCGAATGGATTCGGAGGACGTGTACGATGATGTTGAAACTATACCCATG
AAAAATATGGAAAATGGTAATAATGGTTATCATAAAGATCATGTATATGGTATTCA
CAATCCTGTTATGACAAGCCCATCACAACATTGATAGCTGATGTTGAAACTGCTTG
AGCATCAGGATACTCAAAGTGGAAAGGATCACAGATTTTTGGTAGTTTCTGGGTCT
ACAAGGACTTTCCAAATCCAGGAGCAACGCCAGTGGCAACGTAGTGACTCAGGCGG
GCACCAAGGCAACGGCACCATTGGTCTCTGGGTAGTGCTTTAAGAATGAACACAAT
CACGTTATAGTCCATGGTCCATCACTATTCAAGGATGACTCCCTCCCTTCCTGTCT
ATTTTTGTTTTTTACTTTTTTACACTGAGTTTCTATTTAGACACTACAACATATGG
GGTGTTTGTTCCCATTGGATGCATTTCTATCAAAACTCTATCAAATGTGATGGCTA
GATTCTAACATATTGCCATGTGTGGAGTGTGCTGAACACACACCAGTTTACAGGAA
AGATGCATTTTGTGTACAGTAAACGGTGTATATACCTTTTGTTACCACAGAGTTTT
TTAAACAAATGAGTATTATAGGACTTTCTTCTAAATGAGCTAAATAAGTCACCATT
GACTTCTTGGTGCTGTTGAAAATAATCCATTTTCACTAAAAGTGTGTGAAACCTAC
AGCATATTCTTCACGCAGAGATTTTCATCTATTATACTTTATCAAAGATTGGCCAT
GTTCCACTTGGAAATGGCATGCAAAAGCCATCATAGAGAAACCTGCGTAACTCCAT
CTGACAAATTCAAAAGAGAGAGAGAGATCTTGAGAGAGAAATGCTGTTCGTTCAAA
AGTGGAGTTGTTTTAACAGATGCCAATTACGGTGTACAGTTTAACAGAGTTTTCTG
TTGCATTAGGATAAACATTAATTGGAGTGCAGCTAACATGAGTATCATCAGACTAG
TATCAAGTGTTCTAAAATGAAATATGAGAAGATCCTGTCACAATTCTTAGATCTGG
TGTCCAGCATGGATGAAACCTTTGAGTTTGGTCCCTAAATTTGCATGAAAGCACAA
GGTAAATATTCATTTGCTTCAGGAGTTTCATGTTGGATCTGTCATTATCAAAAGTG
ATCAGCAATGAAGAACTGGTCGGACAAAATTTAACGTTGATGTAATGGAATTCCAG
ATGTAGGCATTCCCCCCAGGTCTTTTCATGTGCAGATTGCAGTTCTGATTCATTTG
AATAAAAAGGAACTTGG

The NOV5 nucleic acid was identified on chromosome 4 and has 2874 of 2882 bases (99%) identical to a Homo sapiens prominin (mouse)-like 1 (PROML1), mRNA of 3794 nucleotides (GENBANK-ID: gi|5174386|ref|NM_006017.1|) (E = 0.0)

A disclosed NOV5 polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 is 865 amino acid residues and is presented using the one-letter code in Table 5B. Signal P, Psort and/or Hydropathy results predict that NOV5 has is likely to be localized in the plasma membrane.

Table 5B. Encoded NOV5 protein sequence (SEQ ID NO:10)

```
MALVLGSLLLLGLCGNSFSGGQPSSTDAPKAWNYELPATNYETQDSHKAGPIGILFELVHIFLYV
VQPRDFPEDTLRKFLQKAYESKIDYDKPETVILGLKIVYYEAGIILCCVLGLLFIILMPLVGYFF
CMCRCCNKCGGEMHQRQKENGPFLRKCFAISLLVICIIISIGIFYGFVANHQVRTRIKRSRKLAD
SNFKDLRTLLNETPEQIKYILAQYNTTKDKAFTDLNSINSVLGGGILDRLRPNIIPVLDEIKSMA
TAIKETKEALENMNSTLKSLHQQSTQLSSSLTSVKTSLRSSLNDPLCLVHPSSETCNSIRLSLSQ
LNSNPELRQLPPVDAELDNVNNVLRTDLDGLVQQGYQSLNDIPDRVQRQTTTVVAGIKRVLNSIG
SDIDNVTQRLPIQDILSAFSVYVNNTESYIHRNLPTLEEYDSYWWLGGLVICSLLTLIVIFYYLG
LLCGVCGYDRHATPTTRGCVSNTGGVFLMVGVGLSFLFCWILMIIVVLTFVFGANVEKLICEPYT
SKELFRVLDTPYLLNEDWEYYLSGKLFNKSKMKLTFEQVYSDCKKNRGTYGTLHLQNSFNISEHL
NINEHTGSISSELESLKVNLNIFLLGAAGRKNLQDFAACGIDRMNYDSYLAQTGKSPAGVNLLSF
AYDLEAKANSLPPGNLRNSLKRDAQTIKTIHQQRVLPIEQSLSTLYQSVKILQRTGNGLLERVTR
ILASLDFAQNFITNNTSSVIIEETKKYGRTIIGYFEHYLQWIEFSISEKVASCKPVATALDTAVD
VFLCSYIIDPLNLFWFGIGKATVFLLPALIFAVKLAKYYRRMDSEDVYDDVETIPMKNMENGNNG
YHKDHVYGIHNPVMTSPSQH
```

The disclosed NOV5 amino acid sequence has 865 of 865 amino acid residues (100%) identical to, and 865 of 864 amino acid residues (100%) similar to, the 865 amino acid residue AC133 antigen from *Homo sapiens* (Human) (GenBank Acc. No.: AF027208) (E = 0.0).

NOV5 is expressed in at least the following tissues: fetal heart, pooled human melanocyte, fetal heart, and pregnant uterus. TaqMan data for NOV5 is shown below in Example 1, and SAGE data is shown below in Example 2. The TaqMan data shows overexpression in cell lines derived from colon, ovarian, lung and liver tumors. The EST analysis showed that NOV5 was found in well-differentiated endometrial adenocarcinoma, 7 pooled tumors, and retina.

NOV5 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 5C.

Table 5C. BLAST results for NOV5

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11437151|ref|XP_003591.1|	prominin (mouse)-like 1 [Homo sapiens]	727	437/480 (91%)	670/718 (93%)	0.0

gi|15082356|gb|AAH12089.1|AAH12089 (BC012089)	Similar to prominin (mouse)-like 1 [Homo sapiens]	856	788/844 (93%)	788/844 (93%)	0.0
gi|5174387|ref|NP_006008.1|	prominin (mouse)-like 1; hematopoietic stem cell antigen [Homo sapiens]	865	797/844 (94%)	797/844 (94%)	0.0
gi|15042603|gb|AAK82364.1|AF386758_1 (AF386758)	prominin [Rattus norvegicus]	857	484/845 (57%)	625/845 (73%),	0.0
gi|13124464|sp|O54990|PROM_MOUSE	PROMININ PRECURSOR (ANTIGEN AC133 HOMOLOG)	867	485/846 (57%)	627/846 (73%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 5D.

Table 5D Clustal W Sequence Alignment

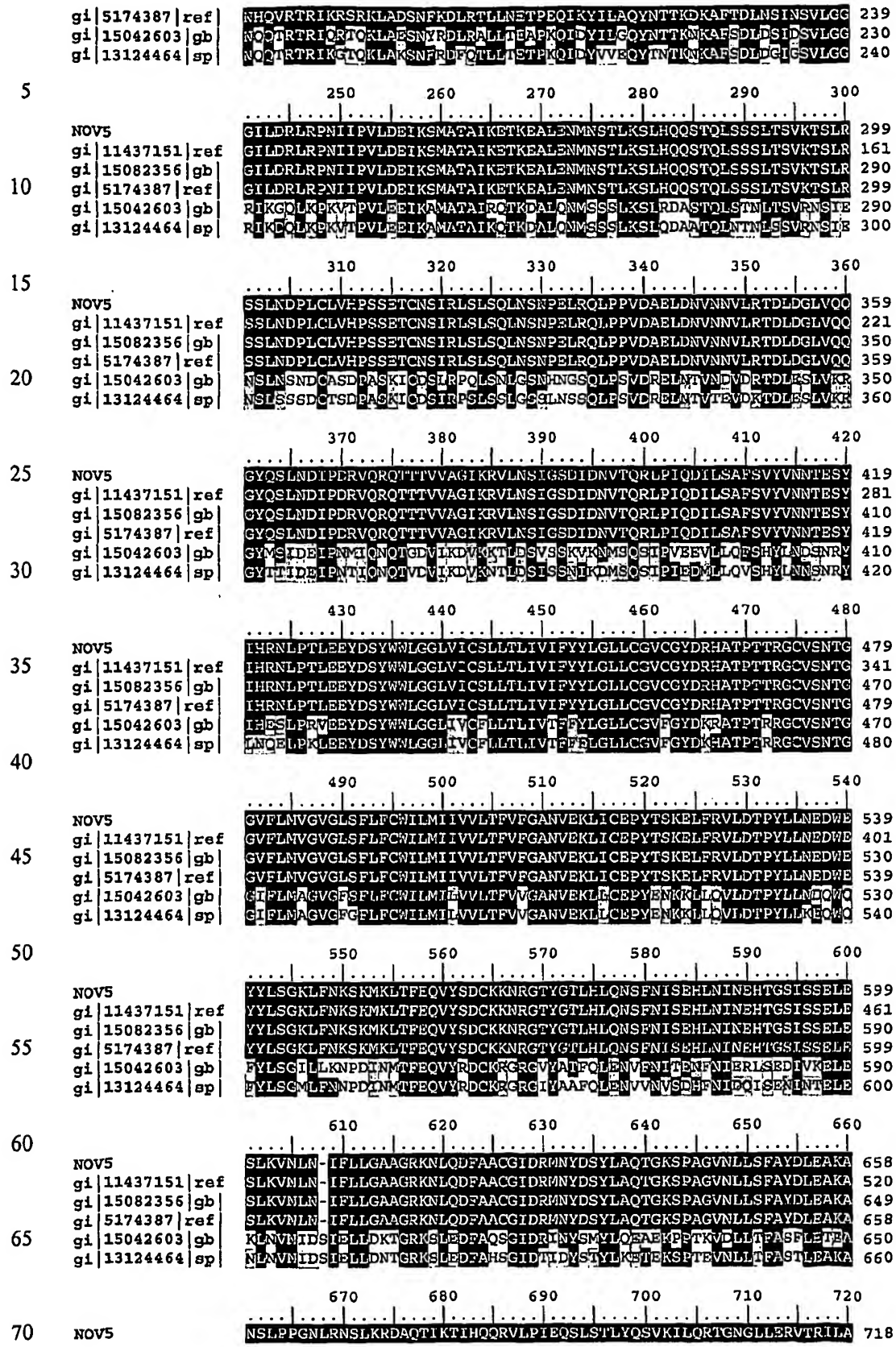
1) NOV5 (SEQ ID NO:10)
2) gi|11437151|ref|XP_003591.1| prominin (mouse)-like 1 [Homo sapiens] (SEQ ID NO:33)
3) gi|15082356|gb|AAH12089.1|AAH12089 (BC012089) Similar to prominin (mouse)-like 1 [Homo sapiens] (SEQ ID NO:34)
4) gi|5174387|ref|NP_006008.1| prominin (mouse)-like 1; hematopoietic stem cell antigen [Homo sapiens] (SEQ ID NO:35)
5) gi|15042603|gb|AAK82364.1|AF386758_1 (AF386758) prominin [Rattus norvegicus] (SEQ ID NO:36
6) gi|13124464|sp|O54990|PROM_MOUSE PROMININ PRECURSOR (ANTIGEN AC133 HOMOLOG) (SEQ ID NO:37)

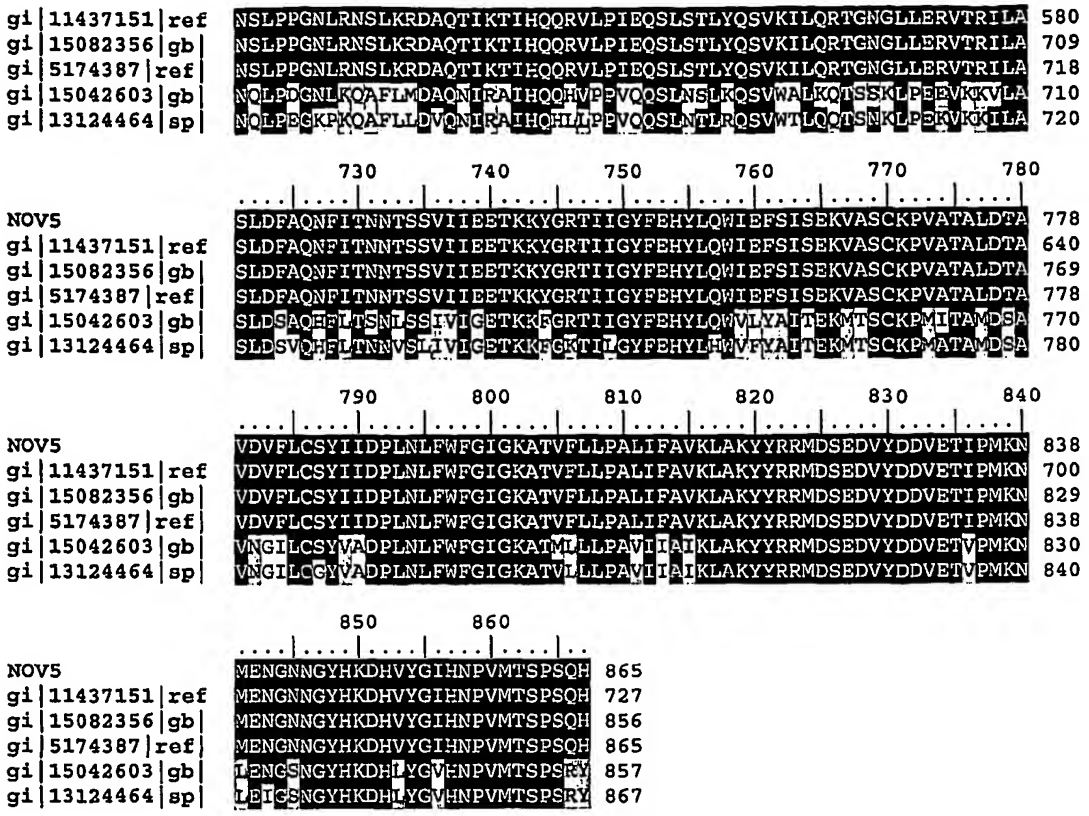
```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
NOV5                    MALVLGSLLLLGLCGNSFSGGQPSSTDAPKAWNYELPATNYETQDSHKAGPIGILFFELVH 60
gi|11437151|ref         ------------------------------------------------------------ 1
gi|15082356|gb|         MALVLGSLLLLGLCGNSFSGGQPSSTDAPKAWNYELPATNYETQDSHKAGPIGILFFELVH 60
gi|5174387|ref|         MALVLGSLLLLGLCGNSFSGGQPSSTDAPKAWNYELPATNYETQDSHKAGPIGILFFELVH 60
gi|15042603|gb|         MALVFSVLLLLGLCGKMASGGQPAFDNTPGALNYELPTTEYETQDTFNAGIIDPLYQMVH 60
gi|13124464|sp|         MALVFSALLLLGLCGKISSEGQPAFHNTPGAMNYELPTTKYETQDTFNAGIVGPLYKMVH 60

                        70        80        90        100       110       120
                        |         |         |         |         |         |
NOV5                    IFLYVVQPRDFPEDTLRKFLQ-KAYESKIDYDKPETVILGLKIVYYEAGIILCCVLGLLF 119
gi|11437151|ref         -----MWWRNAP-------------------------------AT------------ 9
gi|15082356|gb|         IFLYVVQPRDFPEDTLRKFLQ-KAYESKIDYDK---------IVYYEAGIILCCVLGLLF 110
gi|5174387|ref|         IFLYVVQPRDFPEDTLRKFLQ-KAYESKIDYDKPETVILGLKIVYYEAGIILCCVLGLLF 119
gi|15042603|gb|         IFLNVVQPNDFPQDLVKKLIQ-KRFDISVDTKE---------VAIYEIGVLICVILGLLF 110
gi|13124464|sp|         IFLSVVQPNDFPLDLIKKLIQNKKFDISVDSKEPEIIVLALKIALYEIGVLICAILGLLF 120

                        130       140       150       160       170       180
                        |         |         |         |         |         |
NOV5                    IILMPLVGYFFCMCRCCNKCGGEMHQRQKENGPFLRKCFAISLLVICIIISIGIFYGFVA 179
gi|11437151|ref         ------------------------PENGPFLRKCFAISLLVICIIISIGIFYGFVA 41
gi|15082356|gb|         IILMPLVGYFFCMCRCCNKCGGEMHQRQKENGPFLRKCFAISLLVICIIISIGIFYGFVA 170
gi|5174387|ref|         IILMPLVGYFFCMCRCCNKCGGEMHQRQKENGPFLRKCFAISLLVICIIISIGIFYGFVA 179
gi|15042603|gb|         IFLMPLVGFFFCMCRCCNKCGGEMHQRQKCNESCRRKCLAISLLLICLLMSLGIAFGFVA 170
gi|13124464|sp|         IILMPLVGCFFCMCRCCNKCGGEMHQRQKCNAPCRRKCLGISLLVICLLMSLGIIYGFVA 180

                        190       200       210       220       230       240
                        |         |         |         |         |         |
NOV5                    NHQVRTRIKRSRKLADSNFKDLRTLLNETPEQIKYILAQYNTTKDKAFTDLNSINSVLGG 239
gi|11437151|ref         NHQVRTRIKRSRKLADSNFKDLRTLLNETPEQIKYILAQYNTTKDKAFTDLNSINSVLGG 101
gi|15082356|gb|         NHQVRTRIKRSRKLADSNFKDLRTLLNETPEQIKYILAQYNTTKDKAFTDLNSINSVLGG 230
```


gi|5174387|ref|
gi|15042603|gb|
gi|13124464|sp|
NOV5
gi|11437151|ref
gi|15082356|gb|
gi|5174387|ref|
gi|15042603|gb|
gi|13124464|sp|
250 260 270 280 290 300
310 320 330 340 350 360
370 380 390 400 410 420
430 440 450 460 470 480
490 500 510 520 530 540
550 560 570 580 590 600
610 620 630 640 650 660
670 680 690 700 710 720



MoAb AC133 is an antibody with specificity for a novel cell surface antigen that is expressed on CD34bright subpopulations of HSCs found in adult bone marrow, fetal bone marrow and liver, cord blood, and adult peripheral blood. MoAb AC133 can be used for magnetic bead immunoselection of HSC populations for transplantation, as well as for phenotypic analysis of stem and progenitor cell populations using flow cytometric techniques. The AC133 antigen is a glycosylated protein with a molecular weight of 120 kD.The AC133 polypeptide has a predicted size of 97 kD and contains five transmembrane (5-TM) domains with an extracellular N-terminus and a cytoplasmic C-terminusm (containing 5 tyrosine residues, potential for signalling), 2 small cysteine-rich cytoplasmic loops, and 2 very large extracellular loops each containing 4 consensus sequences for N-linked glycosylation.

The AC133 antigen transcript was also noted in nonlymphoid tissues, particularly the pancreas, kidney, and placenta. Weaker signals were observed for the liver, lung, brain, and heart. This is in contrast to immunohistochemical staining of paraffin tissue sections, where AC133 antigen expression was detectable only in bone marrow. Its presence on early, undifferentiated cells is suggestive of a growth factor receptor, and the presence of five tyrosine residues on the 50-aa cytoplasmic tail may indicate that the protein is phosphorylated in response to ligand binding and initiates a signal transduction cascade. (Miraglia S, Godfrey

W, Yin AH, Atkins K, Warnke R, Holden JT, Bray RA, Waller EK, Buck DW) A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. Blood. 1997 Dec 15;90(12):5013-21.) Human CD34+ progenitor cells expressed AC133, expression was rapidly downregulated during differentiation. In apparent contrast to normal primitive haematopoietic cells, the AC133 protein was undetectable on cells from 24 different human haematopoietic cells lines, even though the majority of these cells expressed AC133 mRNA. (Majka M, Ratajczak J, Machalinski B, Carter A, Pizzini D, Wasik MA, Gewirtz AM, Ratajczak MZ). Expression, regulation and function of AC133, a putative cell surface marker of primitive human haematopoietic cells. (Folia Histochem Cytobiol. 2000;38(2):53-63.)

The human AC133 antigen and mouse prominin are structurally related plasma membrane proteins. The human AC133 antigen shows the features characteristic of mouse prominin in epithelial and transfected non-epithelial cells, i.e. a selective association with apical microvilli and plasma membrane protrusions, respectively. Conversely, flow cytometry of murine CD34(+) bone marrow progenitors revealed the cell surface expression of prominin. Taken together, the data strongly suggest that the AC133 antigen is the human orthologue of prominin. (Corbeil D, Roper K, Hellwig A, Tavian M, Miraglia S, Watt SM, Simmons PJ, Peault B, Buck DW, Huttner WB). The human AC133 hematopoietic stem cell antigen is also expressed in epithelial cells and targeted to plasma membrane protrusions. (J Biol Chem. 2000 Feb 25;275(8):5512-20.)

NOV5 is thought to be involved in metastatic potential and chemotherapy resistance. Therapeutic targeting of AC133 with a monoclonal antibody is anticipated to limit or block the extent of metastasis and chemotherapy resistance in colon, gastric, ovarian and lung tumors.

The disclosed NOV5 nucleic acid of the invention encoding a AC133 Antigen -like protein includes the nucleic acid whose sequence is provided in Table 5A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 5A while still encoding a protein that maintains its AC133 Antigen -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or

derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV5 protein of the invention includes the AC133 Antigen -like protein whose sequence is provided in Table 5B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 5B while still encoding a protein that maintains its AC133 Antigen -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 43% percent of the residues may be so changed.

NOV5 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders. These antibodies could also be used to treat certain pathologies as described above.

NOV6

A disclosed NOV6 nucleic acid of 1807 nucleotides (also referred to as NM_012445) encoding a novel Spondin 2-like protein is shown in Table 6A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 276-278 and ending with a TAA codon at nucleotides 1269-1271. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 6A, and the start and stop codons are in bold letters.

Table 6A. NOV6 Nucleotide Sequence (SEQ ID NO:11)

GCACGAGGGAAGAGGGTGATCCGACCCGGGGAAGGTCGCTGGGCAGGGCGAGTTGGGAAAGCGGCAGCCC
CCGCCGCCCCCGCAGCCCCTTCTCCTCCTTTCTCCCACGTCCTATCTGCCTCTCGCTGGAGGCCAGGCCG
TGCAGCATCGAAGACAGGAGGAACTGGAGCCTCATTGGCCGGCCCGGGGCGCCGGCCTCGGGCTTAAATA
GGAGCTCCGGGGCTCTGGCTGGGACCCGACCGCTGCCGGCCGCGCTCCCGCTGCTCCTGCCGGGTGATGGA
AAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTCTGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCC
GGCCAGCCTCTTGGGGGAGAGTCCATCTGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGG
GCAAGTGGAGCCAGACGGCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCT
GCTGGGGGCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTGCGC
GACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGGGAGGCGCTGCAGA
GCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGGCAGACGTCGGCGGAGCTGGAGGT
GCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTG
GACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACG

CCGGGACGGACAGCGGCTTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGA
GATAACGTCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCTCCC
ATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGC
CCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCT
GTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGACCAAGAGCAGGACTCGCTACGTC
CGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATA
ACTGCGTCTAAGACCAGAGCCCCGCAGCCCCTGGGGCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTG
CAGGCTCATGCTGCAGGCGGCCGAGGCACAGGGGGTTTCGCGCTGCTCCTGACCGCGGTGAGGCCGCGCC
GACCATCTCTGCACTGAAGGGCCCTCTGGTGGCCGGCACGGGCATTGGGAAACAGCCTCCTCCTTTCCCA
ACCTTGCTTCTTAGGGGCCCCCGTGTCCCGTCTGCTCTCAGCCTCCTCCTCCTGCAGGATAAAGTCATCC
CCAAGGCTCCAGCTACTCTAAATTATGGTCTCCTTATAAGTTATTGCTGCTCCAGGAGATTGTCCTTCAT
CGTCCAGGGGCCTGGCTCCCACGTGGTTGCAGATACCTCAGACCTGGTGCTCTAGGCTGTGCTGAGCCCA
CTCTCCCGAGGGCGCATCCAAGCGGGGGCCACTTGAGAAGTGAATAAATGGGGCGGTTTCGGAAGCGTCA
GTGTTTCCATGTTATGGATCTCTCTGCGTTTGAATAAAGACTATCTCTGTTGCTCAC

The disclosed NOV6 nucleic acid sequence localized to chromosome 4, has 1587 of 1591 bases (99%) identical to a Homo sapiens spondin 2, extracellular matrix protein (SPON2), mRNA (GENBANK-ID: gi|14728622|ref|XM_042674.1|) (E = 0.0).

A disclosed NOV6 polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 is 331 amino acid residues and is presented using the one-letter amino acid code in Table 6B. Signal P, Psort and/or Hydropathy results predict that NOV6 is likely to be localized extracellularly.

Table 6B. Encoded NOV6 protein sequence (SEQ ID NO:12).

MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGA
AHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVS
FVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSF
YYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS
RTRYVRVQPANNGSPCPELEEEAECVPDNCV

The disclosed NOV6 amino acid sequence has 877 of 879 amino acid residues (99%) identical to, and 878 of 879 amino acid residues (99%) similar to, the 879 amino acid residue SPONDIN 2 3 PROTEIN protein from *Mus musculus* (Mouse (Q9QYS2) (E = 0.0).

TaqMan data for NOV6 is shown below in Example 1. It shows overexpression in selected tumor derived cell lines and liver cancers.

NOV6 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 6C.

Table 6C. BLAST results for NOV6

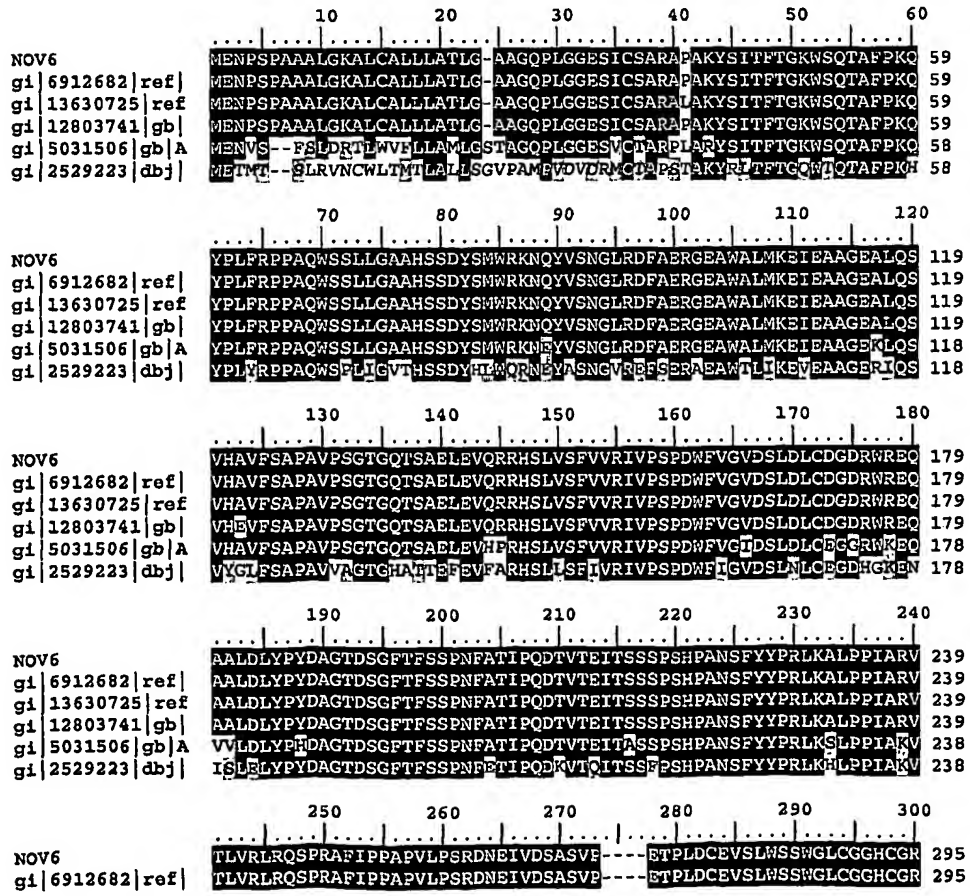
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|6912682|ref|NP_036577.1|	spondin 2, extracellular matrix protein [Homo sapiens]	331	306/331 (92%)	306/331 (92%)	e-163

gi|13630725|ref|XP_003447.2|	spondin 2, extracellular matrix protein [Homo sapiens]	331	305/331 (92%)	305/331 (92%)	e-163
gi|12803741|gb|AAH02707.1|AAH02707 (BC002707)	spondin 2, extracellular matrix protein [Homo sapiens]	331	304/331 (91%)	305/331 (91%)	e-163
gi|5031506|gb|AAD38195.1|AF155196_1 (AF155196)	mindin precursor [Rattus norvegicus]	330	268/300 (89%)	282/300 (93%)	e-149
gi|2529223|dbj|BAA22809.1| (AB006085)	MINDIN2 [Danio rerio]	331	192/304 (63%)	241/304 (79%)	e-113

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 6D.

Table 6D Information for the ClustalW proteins

1) NOV6 (SEQ ID NO:12)
2) gi|6912682|ref|NP_036577.1| spondin 2, extracellular matrix protein [Homo sapiens] (SEQ ID NO:38)
3) gi|13630725|ref|XP_003447.2| spondin 2, extracellular matrix protein [Homo sapiens] (SEQ ID NO:39).
4) gi|12803741|gb|AAH02707.1|AAH02707 (BC002707) spondin 2, extracellular matrix protein [Homo sapiens] (SEQ ID NO:40)
5) gi|5031506|gb|AAD38195.1|AF155196_1 (AF155196) mindin precursor [Rattus norvegicus] (SEQ ID NO:41)
6) gi|2529223|dbj|BAA22809.1| (AB006085) MINDIN2 [Danio rerio] (SEQ ID NO:42)



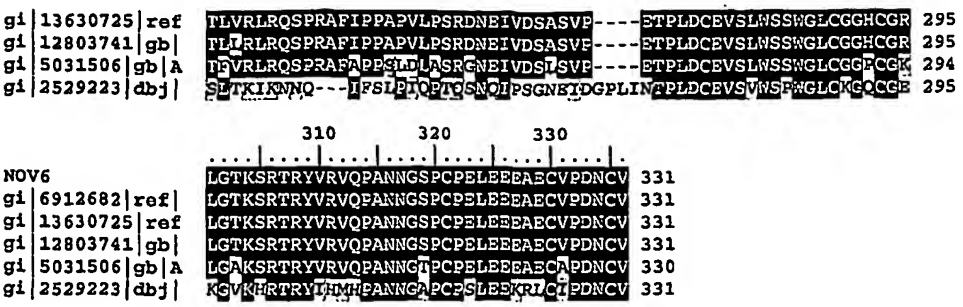


Table 6E-F lists the domain description from DOMAIN analysis results against NOV6. This indicates that the NOV6 sequence has properties similar to those of other proteins known to contain this domain.

Table 6E. Domain Analysis of NOV6

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta (SEQ ID NO:46)
CD-Length = 51 residues, 98.0% aligned
Score = 42.4 bits (98), Expect = 4e-05

```
Query:  280  VSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAE-CVPDNC  330
             || |  |   ||  | ++|||     | | | ||   ++|   |    |
Sbjct:  1    WGEWSEWSPCSVTCGG-GVQTRTRCCN-PPPNGGGPCTGPDTETRACNEQPC  50
```

It is thought that NOV6 is involved with liver cancer. Therapeutic targeting of NOV6 with a monoclonal antibody is anticipated to limit or block the extent of angiogenesis and tumor growth in liver cancer.

The disclosed NOV6 nucleic acid of the invention encoding a Spondin 2 -like protein includes the nucleic acid whose sequence is provided in Table 6A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 6A while still encoding a protein that maintains its Spondin 2 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be

used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV6 protein of the invention includes the Spondin 2 -like protein whose sequence is provided in Table 6B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 6B while still encoding a protein that maintains its Spondin 2 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 37% percent of the residues may be so changed.

NOV6 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV6 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders. These antibodies could also be used to treat certain pathogies as detailed above.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring

polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver,

spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, and 11, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides SEQ ID NOS:1, 3, 5, 7, 9, and 11, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, and 11, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an NOVX

polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, or 11 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, or 11 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, and 11, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of

hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, and 11, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350

or 400 consecutive sense strand nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, or 11; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, or 11; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, and 11.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, or 11, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, and 11 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, and 11. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, or 12.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, and 11, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules

comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:1, 3, 5, 7, 9, and 11 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target

sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:1, 3, 5, 7, 9, and 11, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, and 11, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting

example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.,* Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, and 11, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2, 4, 6, 8, 10, or 12. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:1, 3, 5, 7, 9, and 11 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, and 12. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, and 12; more preferably at least about 70% homologous SEQ ID NOS:2, 4, 6, 8, 10, or 12; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, or 12; even more preferably at least about 90%

homologous to SEQ ID NOS:2, 4, 6, 8, 10, or 12; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, or 12.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, or 12 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:1, 3, 5, 7, 9, and 11 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:1, 3, 5, 7, 9, and 11, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and an NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NOS:2, 4, 6, 8, 10, or 12, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, and 11, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example,

the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylçytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered

systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.,* Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.,* Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.,* Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, and 11). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic

RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking,

number of bonds between the nucleobases, and orientation (*see,* Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, or 12. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, or 12 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed

by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, or 12) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2, 4, 6, 8, 10, or 12. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, or 12, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, or 12, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, or 12, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2, 4, 6, 8, 10, or 12.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings

for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, and 11.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2, 4, 6, 8, 10, or 12, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX

polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR

BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

5

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein).
10 An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific
15 biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics)
20 or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example,
25 enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical
30 synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3;

Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such

antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow and Lane, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson

(The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San

Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., MONOCLONAL ANTIBODY PRODUCTION TECHNIQUES AND APPLICATIONS, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the

immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeyen et al., *Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma

technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al,(*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human

variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol comSlitg agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with

enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation

using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-NOVX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for NOVX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-NOVX antibody (*e.g.*, monoclonal antibody) can be used to isolate an NOVX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-NOVX antibody can facilitate the purification of natural NOVX polypeptide from cells and of recombinantly-produced NOVX polypeptide expressed in host cells. Moreover, an anti-NOVX antibody can be used to detect NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the NOVX protein. Anti-NOVX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include

luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN

ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION

TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43:

235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful

for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NOS:1, 3, 5, 7, 9, and 11 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, and 11), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:1, 3, 5, 7, 9, and 11 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be

used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by

including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries,

while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the

assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule. As used herein, a "target molecule" is a molecule with which an NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An NOVX target molecule can be a non-NOVX molecule or an NOVX protein or polypeptide of the invention. In one embodiment, an NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*,

luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to an NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate an NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent

such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in

the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to

interact, *in vivo*, forming an NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NOS:1, 3, 5, 7, 9, and 11, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding

regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:1, 3, 5, 7, 9, and 11 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in an NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or

prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS:1, 3, 5, 7, 9, and 11, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in*

vitro techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test

sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from an NOVX gene; (*ii*) an addition of one or more nucleotides to an NOVX gene; (*iii*) a substitution of one or more nucleotides of an NOVX gene, (*iv*) a chromosomal rearrangement of an NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of an NOVX gene, (*vi*) aberrant modification of an NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an NOVX gene, (*viii*) a non-wild-type level of an NOVX protein, (*ix*) allelic loss of an NOVX gene, and (*x*) inappropriate post-translational modification of an NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional

means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human*

Mutation 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.*, Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.*, Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85:

4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.*, Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.*, Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.*, Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.*, Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.*, Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.*, Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.*, Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.*, Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.*, Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic

polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting

increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of an NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be

utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, an NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an NOVX protein, a peptide, an NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX

that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering an NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder,

immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

Examples

Example 1. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/μl RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

PANEL 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and

grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 μg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^6$ cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol (5.5 x 10^{-5} M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100

ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 μg/ml or anti-CD40 (Pharmingen) at approximately 10 μg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μg/ml anti-CD28 (Pharmingen) and 2 μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems,

German Town, MD) were cultured at 10^5-10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 □g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 □g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 □g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5 x10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 x10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase

was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 µl of RNAse-free water and 35 µl buffer (Promega) 5 µl DTT, 7 µl RNAsin and 8 µl DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C.

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA

contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

NOV1

Expression of gene NOV1 was assessed using the primer-probe sets Ag1395, described in Table 7. Results from RTQ-PCR runs are shown in Tables 8 and 9.

Table 7. Probe and Primer Ag 1395

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CTGCACTTCAAGGACAGTTACC-3'	59.8	22	2184	50
Probe	FAM-5'-CTATCCATCCACGATGTGCCCAGCT-3'-TAMRA	71.1	25	2217	51
Reverse	5'-TGACAAGGAGCTTACTCTTCCA-3'	59.1	22	2247	52

Table 8. Panel 1.2

Tissue Name	Rel. Expr., % 1.2tm1636f_ag1395	Rel. Expr., % 1.2tm1675f_ag1395*
Endothelial cells	0	0
Heart (fetal)	0.2	0.1
Pancreas	0	0
Pancreatic ca. CAPAN 2	0.4	0.6
Adrenal Gland (new lot*)	1.1	3.6
Thyroid	0	0
Salavary gland	0.2	0.3
Pituitary gland	0	0
Brain (fetal)	1.8	1.9
Brain (whole)	11.3	3.3
Brain (amygdala)	9.8	18.2
Brain (cerebellum)	3.1	3.6
Brain (hippocampus)	31.4	42.6
Brain (thalamus)	2.1	2.9

Cerebral Cortex	100	100
Spinal cord	0.1	0
CNS ca. (glio/astro) U87-MG	0	0
CNS ca. (glio/astro) U-118-MG	0	0
CNS ca. (astro) SW1783	0	0
CNS ca.* (neuro; met) SK-N-AS	0.1	0.3
CNS ca. (astro) SF-539	0	0
CNS ca. (astro) SNB-75	0	0
CNS ca. (glio) SNB-19	0	0
CNS ca. (glio) U251	0	0
CNS ca. (glio) SF-295	0.1	0.1
Heart	0	0.3
Skeletal Muscle (new lot*)	0	0
Bone marrow	0.9	0.8
Thymus	0	0
Spleen	0	0.1
Lymph node	0	0
Colorectal	0	0
Stomach	0.3	0.1
Small intestine	0.2	0.2
Colon ca. SW480	0.5	0.1
Colon ca.* (SW480 met)SW620	0.2	0.1
Colon ca. HT29	0	0
Colon ca. HCT-116	1.3	1.8
Colon ca. CaCo-2	0	0
83219 CC Well to Mod Diff (ODO3866)	0	0
Colon ca. HCC-2998	3.2	3.4
Gastric ca.* (liver met) NCI-N87	0	0
Bladder	0.8	0.8
Trachea	0	0
Kidney	0	0
Kidney (fetal)	0	0
Renal ca. 786-0	0.1	0.1
Renal ca. A498	6	4.7
Renal ca. RXF 393	0	0
Renal ca. ACHN	0.8	1
Renal ca. UO-31	0.3	0.2
Renal ca. TK-10	6	3
Liver	0.3	0.3
Liver (fetal)	0	0.1
Liver ca. (hepatoblast) HepG2	0	0
Lung	0	0
Lung (fetal)	0	0
Lung ca. (small cell)LX-1	0	0
Lung ca. (small cell) NCI-H69	16.3	9.3
Lung ca. (s.cell var.) SHP-77	0.4	0.4
Lung ca. (large cell)NCI-H460	0	0
Lung ca. (non-sm. cell) A549	0	0
Lung ca. (non-s.cell) NCI-H23	0.4	0.4
Lung ca (non-s.cell) HOP-62	0	0

Lung ca. (non-s.cl) NCI-H522	9	11.5
Lung ca. (squam.) SW 900	1.5	0.9
Lung ca. (squam.) NCI-H596	18.8	16.6
Mammary gland	0.1	0.1
Breast ca.* (pl. effusion) MCF-7	0	0.2
Breast ca.* (pl.ef) MDA-MB-231	0	0
Breast ca.* (pl. effusion) T47D	0.5	1.3
Breast ca. BT-549	0	0
Breast ca. MDA-N	0	0
Ovary	0.4	0.3
Ovarian ca. OVCAR-3	0	0
Ovarian ca. OVCAR-4	0.2	0.3
Ovarian ca. OVCAR-5	18.4	11.7
Ovarian ca. OVCAR-8	1	1.4
Ovarian ca. IGROV-1	20.2	11.7
Ovarian ca.* (ascites) SK-OV-3	0.4	0.6
Uterus	0	0
Placenta	0	0
Prostate	0.2	0.2
Prostate ca.* (bone met)PC-3	0	0
Testis	0.2	0
Melanoma Hs688(A).T	0	0
Melanoma* (met) Hs688(B).T	0	0
Melanoma UACC-62	0	0
Melanoma M14	0	0
Melanoma LOX IMVI	0	0
Melanoma* (met) SK-MEL-5	0	0
Adipose	6.5	7

Table 9. Panel 2D

Tissue Name	Rel. Expr., % 2dtm2448f_ag1395	Rel. Expr., % 2dx4tm4720f_ag1395_a2
Normal Colon GENPAK 061003	4.2	1.9
83219 CC Well to Mod Diff (ODO3866)	0.7	1.8
83220 CC NAT (ODO3866)	0	1
83221 CC Gr.2 rectosigmoid (ODO3868)	0	1.1
83222 CC NAT (ODO3868)	0	0
83235 CC Mod Diff (ODO3920)	0	1.2
83236 CC NAT (ODO3920)	0	0.8
83237 CC Gr.2 ascend colon (ODO3921)	0.9	2.3
83238 CC NAT (ODO3921)	0	0.4
83241 CC from Partial Hepatectomy (ODO4309)	0.7	0.2
83242 Liver NAT (ODO4309)	0	0.9
87472 Colon mets to lung (OD04451-01)	0	2.3
87473 Lung NAT (OD04451-02)	0.8	0
Normal Prostate Clontech A+ 6546-1	9	8.2

84140 Prostate Cancer (OD04410)	0	3.5
84141 Prostate NAT (OD04410)	2	1.7
87073 Prostate Cancer (OD04720-01)	0.8	1.7
87074 Prostate NAT (OD04720-02)	0	1.5
Normal Lung GENPAK 061010	3.1	10.9
83239 Lung Met to Muscle (ODO4286)	4.4	4.1
83240 Muscle NAT (ODO4286)	0	0.5
84136 Lung Malignant Cancer (OD03126)	2.2	2.6
84137 Lung NAT (OD03126)	3.2	4.6
84871 Lung Cancer (OD04404)	2.4	1.1
84872 Lung NAT (OD04404)	3.3	4.2
84875 Lung Cancer (OD04565)	0	1.6
84876 Lung NAT (OD04565)	1.7	1.7
85950 Lung Cancer (OD04237-01)	0.8	4.5
85970 Lung NAT (OD04237-02)	3.8	7.1
83255 Ocular Mel Met to Liver (ODO4310)	0	0
83256 Liver NAT (ODO4310)	6.2	2.1
84139 Melanoma Mets to Lung (OD04321)	0.8	0
84138 Lung NAT (OD04321)	3.8	5.3
Normal Kidney GENPAK 061008	0.8	1.6
83786 Kidney Ca, Nuclear grade 2 (OD04338)	1.2	2.8
83787 Kidney NAT (OD04338)	0	1.8
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	5.9	5.4
83789 Kidney NAT (OD04339)	0	0
83790 Kidney Ca, Clear cell type (OD04340)	1.3	7.5
83791 Kidney NAT (OD04340)	0	0.3
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0	2.1
83793 Kidney NAT (OD04348)	0.8	0.8
87474 Kidney Cancer (OD04622-01)	2.2	4.1
87475 Kidney NAT (OD04622-03)	0.7	0.4
85973 Kidney Cancer (OD04450-01)	0	0.4
85974 Kidney NAT (OD04450-03)	0	0
Kidney Cancer Clontech 8120607	27.9	60.6
Kidney NAT Clontech 8120608	0.8	2.1
Kidney Cancer Clontech 8120613	0.8	1.7
Kidney NAT Clontech 8120614	0.7	0.7
Kidney Cancer Clontech 9010320	4.7	6.4
Kidney NAT Clontech 9010321	0	2.7
Normal Uterus GENPAK 061018	0	2.2
Uterus Cancer GENPAK 064011	0	8.9
Normal Thyroid Clontech A+ 6570-1	8.7	1.2
Thyroid Cancer GENPAK 064010	0	0
Thyroid Cancer INVITROGEN A302152	0	2.5
Thyroid NAT INVITROGEN A302153	1.1	0.8
Normal Breast GENPAK 061019	2.8	4.1
84877 Breast Cancer (OD04566)	0	1.8
85975 Breast Cancer (OD04590-01)	28.3	27.5
85976 Breast Cancer Mets (OD04590-03)	13.3	14.2
87070 Breast Cancer Metastasis (OD04655-05)	37.9	100
GENPAK Breast Cancer 064006	12	19.3

Breast Cancer Res. Gen. 1024	33.9	25.2
Breast Cancer Clontech 9100266	6.7	7.7
Breast NAT Clontech 9100265	0.5	9.1
Breast Cancer INVITROGEN A209073	3.7	6.9
Breast NAT INVITROGEN A2090734	0.7	0
Normal Liver GENPAK 061009	0	2.6
Liver Cancer GENPAK 064003	0	1.3
Liver Cancer Research Genetics RNA 1025	0.4	2
Liver Cancer Research Genetics RNA 1026	0	1.6
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	1.6	3.4
Paired Liver Tissue Research Genetics RNA 6004-N	1.4	0.7
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.8	0.8
Paired Liver Tissue Research Genetics RNA 6005-N	0	0
Normal Bladder GENPAK 061001	3.5	3.8
Bladder Cancer Research Genetics RNA 1023	0.8	0.5
Bladder Cancer INVITROGEN A302173	3.2	1.1
87071 Bladder Cancer (OD04718-01)	3.8	2.3
87072 Bladder Normal Adjacent (OD04718-03)	5.2	7.4
Normal Ovary Res. Gen.	3	2.9
Ovarian Cancer GENPAK 064008	3.2	2.9
87492 Ovary Cancer (OD04768-07)	3.5	4.6
87493 Ovary NAT (OD04768-08)	0.9	2.2
Normal Stomach GENPAK 061017	2.7	3.7
Gastric Cancer Clontech 9060358	0.4	0.2
NAT Stomach Clontech 9060359	4.3	1.3
Gastric Cancer Clontech 9060395	3	1.2
NAT Stomach Clontech 9060394	2.5	1
Gastric Cancer Clontech 9060397	100	48
NAT Stomach Clontech 9060396	1	2.2
Gastric Cancer GENPAK 064005	4.9	6.7

NOV2

Expression of gene NOV2 was assessed using the primer-probe sets Ag395 and Ag888, described in Tables 10 and 11. Results from RTQ-PCR runs are shown in Tables 12, 13, 14, 15 and 16.

Table 10. Probe and Primer Ag395

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CAGGAAGAAATAAGCCAAGTCCA-3'		23	1409	53
Probe	TET-5'-TCCTTGGCCTCCCGCCTGC-3'-TAMRA		19	1433	54
Reverse	5'-GAGGTCATGTTCTAGCTTCCCATT-3'		24	1463	55

Table 11. Probe and Primer Ag888

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CATAGCTGACCGCATCTGAA-3'	60	20	3101	56
Probe	FAM-5'-AATGCTCCATCTCCTTGGCTGAGTGA-3'-TAMRA	70.1	26	3130	57
Reverse	5'-GGAGCTAGCATCCATCATCAC-3'	59.7	21	3156	58

Probe and Primer Ag784 mentioned in the provisional application for panel 1 is an error.

Table 12: Panel 1.1 (Ag395)

Tissue Name	Rel. Expr., % tm671t_ag395
Adipose	0.2
Adrenal gland	0.1
Bladder	1.4
Brain (amygdala)	0
Brain (cerebellum)	100
Brain (hippocampus)	0.2
Brain (substantia nigra)	1.2
Brain (thalamus)	0.2
Cerebral Cortex	1.5
Brain (fetal)	0.9
Brain (whole)	4.5
CNS ca. (glio/astro) U-118-MG	0.1
CNS ca. (astro) SF-539	0.2
CNS ca. (astro) SNB-75	0.3
CNS ca. (astro)SW1783	0
CNS ca. (glio) U251	0.1
CNS ca. (glio) SF-295	0.4
CNS ca. (glio) SNB-19	0.1
CNS ca. (glio/astro) U87-MG	0.8
CNS ca.* (neuro; met) SK-N-AS	1.2
Mammary gland	1.4
Breast ca. BT-549	0.2
Breast ca. MDA-N	0.7
Breast ca.* (pl. effusion) T47D	0.5
Breast ca.* (pl. effusion) MCF-7	0.3
Breast ca.* (pl.ef) MDA-MB-231	0
Small intestine	0.6
Colorectal	0.2
Colon ca. HT29	0.1

Colon ca. CaCo-2	1
Colon ca. HCT-15	0.3
Colon ca. HCT-116	0.3
Colon ca. HCC-2998	1.1
Colon ca. SW480	0.3
Colon ca.* (SW480 met)SW620	1
Stomach	0.3
Gastric ca.* (liver met) NCI-N87	0.5
Heart	0.4
Fetal Skeletal	0.5
Skeletal muscle	0.8
Endothelial cells	0.2
Heart (fetal)	0
Kidney	0.7
Kidney (fetal)	0.7
Renal ca. 786-0	0
Renal ca. A498	0.3
Renal ca. ACHN	0.3
Renal ca. TK-10	0.5
Renal ca. UO-31	0
Renal ca. RXF 393	0
Liver	0.5
Liver (fetal)	0.5
Liver ca. (hepatoblast) HepG2	0
Lung	0.1
Lung (fetal)	0.2
Lung ca (non-s.cell) HOP-62	1
Lung ca. (large cell)NCI-H460	0.8
Lung ca. (non-s.cell) NCI-H23	0.2
Lung ca. (non-s.cl) NCI-H522	0.7
Lung ca. (non-sm. cell) A549	0.3
Lung ca. (s.cell var.) SHP-77	0.2
Lung ca. (small cell) LX-1	1.2
Lung ca. (small cell) NCI-H69	0.4
Lung ca. (squam.) SW 900	0
Lung ca. (squam.) NCI-H596	0.5
Lymph node	0.3
Spleen	0.1
Thymus	1.1
Ovary	0
Ovarian ca. IGROV-1	0.1
Ovarian ca. OVCAR-3	7.7
Ovarian ca. OVCAR-4	6.4
Ovarian ca. OVCAR-5	1.5
Ovarian ca. OVCAR-8	0.5
Ovarian ca.* (ascites) SK-OV-3	0.7
Pancreas	0.9
Pancreatic ca. CAPAN 2	0
Pituitary gland	0.5
Placenta	0.6

Prostate	2.4
Prostate ca.* (bone met)PC-3	0.2
Salavary gland	2.4
Trachea	1.9
Spinal cord	0.4
Testis	2
Thyroid	0.1
Uterus	0.1
Melanoma M14	0.4
Melanoma LOX IMVI	0.1
Melanoma UACC-62	0
Melanoma SK-MEL-28	1.6
Melanoma* (met) SK-MEL-5	0.1
Melanoma Hs688(A).T	0
Melanoma* (met) Hs688(B).T	0.1

Table 13: Panel 1.2 (Ag888)

Tissue Name	Rel. Expr., % 1.2tm1002f_ag888	Rel. Expr., % 1.2tm1042f_ag888
Endothelial cells	0	0
Heart (fetal)	0	0
Pancreas	0.2	0
Pancreatic ca. CAPAN 2	0	0
Adrenal Gland (new lot*)	0	0
Thyroid	0	0
Salavary gland	8.8	2.7
Pituitary gland	0.5	0
Brain (fetal)	0.7	0
Brain (whole)	22.7	20.2
Brain (amygdala)	0.5	0
Brain (cerebellum)	100	100
Brain (hippocampus)	0.4	0
Brain (thalamus)	0.2	0
Cerebral Cortex	2.7	0
Spinal cord	0.2	0
CNS ca. (glio/astro) U87-MG	0	0
CNS ca. (glio/astro) U-118-MG	0	0
CNS ca. (astro) SW1783	0	0
CNS ca.* (neuro; met) SK-N-AS	0	0
CNS ca. (astro) SF-539	0	0
CNS ca. (astro) SNB-75	0.2	0
CNS ca. (glio) SNB-19	0	0
CNS ca. (glio) U251	0	0
CNS ca. (glio) SF-295	0	0
Heart	0	0
Skeletal Muscle (new lot*)	0	0
Bone marrow	0.3	0

Thymus	0.8	0
Spleen	0	0
Lymph node	0.2	0
Colorectal	0.1	0
Stomach	0.3	0
Small Intestine	0	0
Colon ca. SW480	0	0
Colon ca.* (SW480 met)SW620	0.1	0
Colon ca. HT29	0	0
Colon ca. HCT-116	0	0
Colon ca. CaCo-2	0	0
83219 CC Well to Mod Diff (ODO3866)	0	0
Colon ca. HCC-2998	0	0
Gastric ca.* (liver met) NCI-N87	0	0
Bladder	1.3	0
Trachea	3.7	1.2
Kidney	0.4	0
Kidney (fetal)	1.7	0.2
Renal ca. 786-0	0	0
Renal ca. A498	0.1	0
Renal ca.RXF 393	0	0
Renal ca. ACHN	0	0
Renal ca. UO-31	0	0
Renal ca. TK-10	0	0
Liver	0	0
Liver (fetal)	0	0
Liver ca. (hepatoblast) HepG2	0	0
Lung	0	0
Lung (fetal)	0	0
Lung ca. (small cell) LX-1	0.3	0
Lung ca. (small cell) NCI-H69	1.4	0
Lung ca. (s.cell var.) SHP-77	0	0
Lung ca. (large cell)NCI-H460	0.1	0
Lung ca. (non-sm. cell) A549	0	0
Lung ca. (non-s.cell) NCI-H23	0	0
Lung ca (non-s.cell) HOP-62	0	0
Lung ca. (non-s.cl) NCI-H522	0	0
Lung ca. (squam.) SW 900	0	0
Lung ca. (squam.) NCI-H596	0.7	0
Mammary gland	5.8	2.9
Breast ca.* (pl. effusion) MCF-7	0	0
Breast ca.* (pl.ef) MDA-MB-231	0	0
Breast ca.* (pl. effusion) T47D	0.2	0
Breast ca. BT-549	0	0
Breast ca. MDA-N	0	0
Ovary	0	0
Ovarian ca. OVCAR-3	29.3	18.3
Ovarian ca. OVCAR-4	35.6	22.2
Ovarian ca. OVCAR-5	0.5	0
Ovarian ca.OVCAR-8	0	0

Ovarian ca.IGROV-1	0	0
Ovarian ca.* (ascites) SK-OV-3	0.3	0
Uterus	0	0
Placenta	1.1	0.2
Prostate	3.8	0.6
Prostate ca.* (bone met)PC-3	0	0
Testis	20.6	10.5
Melanoma Hs688(A).T	0	0
Melanoma* (met) Hs688(B).T	0	0
Melanoma UACC-62	0	0
Melanoma M14	0	0
Melanoma LOX IMVI	0	0
Melanoma* (met) SK-MEL-5	0.2	0
Adipose	1.6	0

Table 14. Panel 1.3D (Ag888)

Tissue Name	Rel. Expr., % 1.3dx4tm5629f_ag888_b2
Adipose	0
Adrenal gland	0
Bladder	0
Bone marrow	0
Brain (amygdala)	0.1
Brain (cerebellum)	100
Brain (fetal)	0.1
Brain (hippocampus)	0.2
Cerebral Cortex	0.2
Brain (substantia nigra)	0.4
Brain (thalamus)	0.1
Brain (whole)	19.5
Colorectal	0.1
Heart (fetal)	0
Liver adenocarcinoma	0
Heart	0
Kidney	0.2
Kidney (fetal)	0
Liver	0
Liver (fetal)	0
Lung	0
Lung (fetal)	0
Lymph node	0
Mammary gland	1.2
Fetal Skeletal	0
Ovary	0
Pancreas	0
Pituitary gland	0.3
Placenta	1.4

Prostate	0.6
Salivary gland	1.4
Skeletal muscle	0
Small Intestine	0
Spinal cord	0.1
Spleen	0
Stomach	0.2
Testis	3.5
Thymus	1
Thyroid	0
Trachea	1
Uterus	0
genomic DNA control	93.7
Chemistry Control	67.6

Table 15. Panel 3D (Ag395)

Tissue Name	Rel. Expr., % 2Dtm2317t_ag395
Normal Colon GENPAK 061003	20.2
83219 CC Well to Mod Diff (ODO3866)	6
83220 CC NAT (ODO3866)	5.8
83221 CC Gr.2 rectosigmoid (ODO3868)	1.8
83222 CC NAT (ODO3868)	1.9
83235 CC Mod Diff (ODO3920)	2.2
83236 CC NAT (ODO3920)	5.6
83237 CC Gr.2 ascend colon (ODO3921)	1.2
83238 CC NAT (ODO3921)	0.9
83241 CC from Partial Hepatectomy (ODO4309)	0.9
83242 Liver NAT (ODO4309)	1.3
87472 Colon mets to lung (OD04451-01)	2.2
87473 Lung NAT (OD04451-02)	5.4
Normal Prostate Clontech A+ 6546-1	43.8
84140 Prostate Cancer (OD04410)	17.3
84141 Prostate NAT (OD04410)	15.7
87073 Prostate Cancer (OD04720-01)	41.2
87074 Prostate NAT (OD04720-02)	22.8
Normal Lung GENPAK 061010	2.8
83239 Lung Met to Muscle (ODO4286)	0
83240 Muscle NAT (ODO4286)	66
84136 Lung Malignant Cancer (OD03126)	3.5
84137 Lung NAT (OD03126)	2.9
84871 Lung Cancer (OD04404)	46
84872 Lung NAT (OD04404)	16.6
84875 Lung Cancer (OD04565)	100
84876 Lung NAT (OD04565)	3
85950 Lung Cancer (OD04237-01)	2.6
85970 Lung NAT (OD04237-02)	0.6
83255 Ocular Mel Met to Liver (ODO4310)	1

83256 Liver NAT (ODO4310)	0
84139 Melanoma Mets to Lung (OD04321)	3.5
84138 Lung NAT (OD04321)	0.8
Normal Kidney GENPAK 061008	11.3
83786 Kidney Ca, Nuclear grade 2 (OD04338)	6.2
83787 Kidney NAT (OD04338)	3.6
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	23.8
83789 Kidney NAT (OD04339)	15
83790 Kidney Ca, Clear cell type (OD04340)	3.2
83791 Kidney NAT (OD04340)	11.9
83792 Kidney Ca, Nuclear grade 3 (OD04348)	1.3
83793 Kidney NAT (OD04348)	12.2
87474 Kidney Cancer (OD04622-01)	4.9
87475 Kidney NAT (OD04622-03)	3.1
85973 Kidney Cancer (OD04450-01)	0.5
85974 Kidney NAT (OD04450-03)	7.4
Kidney Cancer Clontech 8120607	3
Kidney NAT Clontech 8120608	1.1
Kidney Cancer Clontech 8120613	0.9
Kidney NAT Clontech 8120614	2
Kidney Cancer Clontech 9010320	13.1
Kidney NAT Clontech 9010321	11.5
Normal Uterus GENPAK 061018	2.9
Uterus Cancer GENPAK 064011	21.3
Normal Thyroid Clontech A+ 6570-1	0.8
Thyroid Cancer GENPAK 064010	2.5
Thyroid Cancer INVITROGEN A302152	3
Thyroid NAT INVITROGEN A302153	0
Normal Breast GENPAK 061019	44.1
84877 Breast Cancer (OD04566)	5.3
85975 Breast Cancer (OD04590-01)	10.8
85976 Breast Cancer Mets (OD04590-03)	6.4
87070 Breast Cancer Metastasis (OD04655-05)	1.4
GENPAK Breast Cancer 064006	13.1
Breast Cancer Res. Gen. 1024	62
Breast Cancer Clontech 9100266	10
Breast NAT Clontech 9100265	12.9
Breast Cancer INVITROGEN A209073	25.2
Breast NAT INVITROGEN A2090734	61.1
Normal Liver GENPAK 061009	5.4
Liver Cancer GENPAK 064003	2.6
Liver Cancer Research Genetics RNA 1025	1
Liver Cancer Research Genetics RNA 1026	0.9
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	9.7
Paired Liver Tissue Research Genetics RNA 6004-N	3.1
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0
Paired Liver Tissue Research Genetics RNA 6005-N	0
Normal Bladder GENPAK 061001	9
Bladder Cancer Research Genetics RNA 1023	2.4
Bladder Cancer INVITROGEN A302173	21.8

87071 Bladder Cancer (OD04718-01)	46.7
87072 Bladder Normal Adjacent (OD04718-03)	4.1
Normal Ovary Res. Gen.	0
Ovarian Cancer GENPAK 064008	65.1
87492 Ovary Cancer (OD04768-07)	33
87493 Ovary NAT (OD04768-08)	0
Normal Stomach GENPAK 061017	2.4
Gastric Cancer Clontech 9060358	1.5
NAT Stomach Clontech 9060359	1.4
Gastric Cancer Clontech 9060395	2.3
NAT Stomach Clontech 9060394	0.8
Gastric Cancer Clontech 9060397	6.6
NAT Stomach Clontech 9060396	0
Gastric Cancer GENPAK 064005	4.5

Table 16. Panel 2D (Ag888)

Tissue Name	Rel. Expr., % 2dtm2313f_ag888	Rel. Expr., % 2Dtm2409f_ag888
Normal Colon GENPAK 061003	10.7	5.6
83219 CC Well to Mod Diff (ODO3866)	0.5	0.5
83220 CC NAT (ODO3866)	0	0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.7	0.2
83222 CC NAT (ODO3868)	0.6	0.7
83235 CC Mod Diff (ODO3920)	2	0.7
83236 CC NAT (ODO3920)	1.1	1.1
83237 CC Gr.2 ascend colon (ODO3921)	0.3	0
83238 CC NAT (ODO3921)	0.8	0.9
83241 CC from Partial Hepatectomy (ODO4309)	0.7	0.2
83242 Liver NAT (ODO4309)	0.9	0
87472 Colon mets to lung (OD04451-01)	0.7	0.4
87473 Lung NAT (OD04451-02)	0.6	0.2
Normal Prostate Clontech A+ 6546-1	29.3	21
84140 Prostate Cancer (OD04410)	9.3	5.2
84141 Prostate NAT (OD04410)	8.9	12.2
87073 Prostate Cancer (OD04720-01)	37.9	41.2
87074 Prostate NAT (OD04720-02)	37.1	33.2
Normal Lung GENPAK 061010	4.5	3
83239 Lung Met to Muscle (ODO4286)	1.3	1.3
83240 Muscle NAT (ODO4286)	24	16.7
84136 Lung Malignant Cancer (OD03126)	4.4	2.4
84137 Lung NAT (OD03126)	1.8	0.2
84871 Lung Cancer (OD04404)	100	30.4
84872 Lung NAT (OD04404)	5.9	1.7
84875 Lung Cancer (OD04565)	65.5	100
84876 Lung NAT (OD04565)	0.8	2
85950 Lung Cancer (OD04237-01)	0.9	1.2
85970 Lung NAT (OD04237-02)	0.9	0.2

83255 Ocular Mel Met to Liver (ODO4310)	0.7	0.9
83256 Liver NAT (ODO4310)	0	0
84139 Melanoma Mets to Lung (OD04321)	1.1	0.3
84138 Lung NAT (OD04321)	1.2	0.5
Normal Kidney GENPAK 061008	10.3	3
83786 Kidney Ca, Nuclear grade 2 (OD04338)	2.3	2.4
83787 Kidney NAT (OD04338)	3.4	1.4
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	4.4	3.3
83789 Kidney NAT (OD04339)	5.8	5.5
83790 Kidney Ca, Clear cell type (OD04340)	1	2.4
83791 Kidney NAT (OD04340)	9.8	4
83792 Kidney Ca, Nuclear grade 3 (OD04348)	2	1.9
83793 Kidney NAT (OD04348)	2.7	3
87474 Kidney Cancer (OD04622-01)	2.5	4.8
87475 Kidney NAT (OD04622-03)	3.2	4.4
85973 Kidney Cancer (OD04450-01)	1.6	0.1
85974 Kidney NAT (OD04450-03)	3	0.8
Kidney Cancer Clontech 8120607	2.7	0.4
Kidney NAT Clontech 8120608	0.4	1.9
Kidney Cancer Clontech 8120613	0	0.4
Kidney NAT Clontech 8120614	0	0.3
Kidney Cancer Clontech 9010320	2.4	0.9
Kidney NAT Clontech 9010321	2.3	3.1
Normal Uterus GENPAK 061018	0.1	1.1
Uterus Cancer GENPAK 064011	23.2	21.2
Normal Thyroid Clontech A+ 6570-1	0.7	0.7
Thyroid Cancer GENPAK 064010	3.2	1.5
Thyroid Cancer INVITROGEN A302152	0.7	1.5
Thyroid NAT INVITROGEN A302153	0.4	0.5
Normal Breast GENPAK 061019	9.2	16.6
84877 Breast Cancer (OD04566)	1.7	3.8
85975 Breast Cancer (OD04590-01)	1.2	1.2
85976 Breast Cancer Mets (OD04590-03)	3.1	3.3
87070 Breast Cancer Metastasis (OD04655-05)	0.2	1.4
GENPAK Breast Cancer 064006	13.7	5.5
Breast Cancer Res. Gen. 1024	55.9	23.3
Breast Cancer Clontech 9100266	22.4	14.4
Breast NAT Clontech 9100265	36.6	28.5
Breast Cancer INVITROGEN A209073	43.8	44.8
Breast NAT INVITROGEN A2090734	100	20.7
Normal Liver GENPAK 061009	0	0.4
Liver Cancer GENPAK 064003	1	0
Liver Cancer Research Genetics RNA 1025	0.4	0.3
Liver Cancer Research Genetics RNA 1026	0	0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0	0
Paired Liver Tissue Research Genetics RNA 6004-N	0.6	0.2
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.3	0.3
Paired Liver Tissue Research Genetics RNA 6005-N	0	0
Normal Bladder GENPAK 061001	2.5	3.1
Bladder Cancer Research Genetics RNA 1023	0.4	0.3

Bladder Cancer INVITROGEN A302173	33.4	11.9
87071 Bladder Cancer (OD04718-01)	75.3	68.3
87072 Bladder Normal Adjacent (OD04718-03)	1.6	0.5
Normal Ovary Res. Gen.	0.4	0
Ovarian Cancer GENPAK 064008	91.4	50.3
87492 Ovary Cancer (OD04768-07)	17.9	10.8
87493 Ovary NAT (OD04768-08)	0	0.2
Normal Stomach GENPAK 061017	2.1	1.6
Gastric Cancer Clontech 9060358	0.7	0
NAT Stomach Clontech 9060359	0.4	0.4
Gastric Cancer Clontech 9060395	0.4	0.2
NAT Stomach Clontech 9060394	0.3	0.7
Gastric Cancer Clontech 9060397	2.8	0.8
NAT Stomach Clontech 9060396	0	0.2
Gastric Cancer GENPAK 064005	1.5	0.3

NOV3

Expression of gene NOV3 was assessed using the primer-probe set Ag784, described in Table 17. Results from RTQ-PCR runs are shown in Tables 12, 13, 14, 15 and 16.

Table 17. Probe and Primer Ag784

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GTCCTGGGATGTGTGAGAGAT-3'	59	21	1147	59
Probe	FAM-5'-CAGAGAGACGCAGCTCCTCCAAGAAG-3'-TAMRA	69.8	26	1174	60
Reverse	5'-GAACAACCTCACAGAGCTTCAG-3'	59.1	22	1223	61

Table 18. Panel 1.2

Tissue Name	Rel. Expr., % 1.2tm924f_ag784	Rel. Expr., % 1.2tm1115f_ag784
Endothelial cells	0	0
Heart (fetal)	0.4	13.1
Pancreas	0	0
Pancreatic ca. CAPAN 2	7.3	0
Adrenal Gland (new lot*)	0	0
Thyroid	22.5	0
Salavary gland	15.2	15.6
Pituitary gland	100	14
Brain (fetal)	2.5	0
Brain (whole)	11.3	0
Brain (amygdala)	0	0
Brain (cerebellum)	20.4	26.2

Brain (hippocampus)	0.3	0
Brain (thalamus)	3.6	0
Cerebral Cortex	0.2	0
Spinal cord	0	0
CNS ca. (glio/astro) U87-MG	0	0
CNS ca. (glio/astro) U-118-MG	0	0
CNS ca. (astro) SW1783	0	0
CNS ca.* (neuro; met) SK-N-AS	0	0
CNS ca. (astro) SF-539	0	0
CNS ca. (astro) SNB-75	0	0
CNS ca. (glio) SNB-19	0	0
CNS ca. (glio) U251	0	0
CNS ca. (glio) SF-295	0	0
Heart	5.4	2.1
Skeletal Muscle (new lot*)	0	0
Bone marrow	0	0
Thymus	0	0
Spleen	5.7	0
Lymph node	0	0
Colorectal	0	1.3
Stomach	0	0
Small intestine	0	0
Colon ca. SW480	19.1	18.7
Colon ca.* (SW480 met)SW620	56.6	8.5
Colon ca. HT29	0	0
Colon ca. HCT-116	0	0
Colon ca. CaCo-2	0	0
83219 CC Well to Mod Diff (ODO3866)	0	0.9
Colon ca. HCC-2998	1.6	0
Gastric ca.* (liver met) NCI-N87	20.7	21.3
Bladder	0	0
Trachea	9.7	11.3
Kidney	0	0
Kidney (fetal)	0	0
Renal ca. 786-0	0	0
Renal ca. A498	0	0
Renal ca. RXF 393	0	0
Renal ca. ACHN	0	0
Renal ca. UO-31	0	0
Renal ca. TK-10	0	0
Liver	0	0
Liver (fetal)	0	0
Liver ca. (hepatoblast) HepG2	0	0
Lung	1.2	0
Lung (fetal)	0	0
Lung ca. (small cell) LX-1	45.4	20.9
Lung ca. (small cell) NCI-H69	28.1	55.9
Lung ca. (s.cell var.) SHP-77	0	0
Lung ca. (large cell)NCI-H460	0	0
Lung ca. (non-sm. cell) A549	27.4	49

Lung ca. (non-s.cell) NCI-H23	0	0
Lung ca (non-s.cell) HOP-62	0	0
Lung ca. (non-s.cl) NCI-H522	0	0
Lung ca. (squam.) SW 900	6.4	0.5
Lung ca. (squam.) NCI-H596	64.6	100
Mammary gland	16	19.6
Breast ca.* (pl. effusion) MCF-7	0	0
Breast ca.* (pl.ef) MDA-MB-231	0	0
Breast ca.* (pl. effusion) T47D	0	0
Breast ca. BT-549	0	0
Breast ca.MDA-N	0	0
Ovary	0	0
Ovarian ca.OVCAR-3	0.2	0
Ovarian ca. OVCAR-4	0	0
Ovarian ca. OVCAR-5	0.6	0
Ovarian ca. OVCAR-8	0	0
Ovarian ca. IGROV-1	0	0
Ovarian ca.* (ascites) SK-OV-3	1	0
Uterus	0	0
Placenta	0	0
Prostate	2.3	7.7
Prostate ca.* (bone met)PC-3	0	0
Testis	0	0
Melanoma Hs688(A).T	0	0
Melanoma* (met) Hs688(B).T	0	0
Melanoma UACC-62	0	0
Melanoma M14	0	0
Melanoma LOX IMVI	0	0
Melanoma* (met) SK-MEL-5	0	0
Adipose	0	0

Table 19. Panel 2D

Tissue Name	Rel. Expr., % 2dtm2311f_ag784
Normal Colon GENPAK 061003	23.8
83219 CC Well to Mod Diff (ODO3866)	22.1
83220 CC NAT (ODO3866)	12.5
83221 CC Gr.2 rectosigmoid (ODO3868)	12
83222 CC NAT (ODO3868)	1.7
83235 CC Mod Diff (ODO3920)	8.1
83236 CC NAT (ODO3920)	9
83237 CC Gr.2 ascend colon (ODO3921)	3.1
83238 CC NAT (ODO3921)	1.3
83241 CC from Partial Hepatectomy (ODO4309)	69.7
83242 Liver NAT (ODO4309)	4.5
87472 Colon mets to lung (OD04451-01)	21.2
87473 Lung NAT (OD04451-02)	12.2

Normal Prostate Clontech A+ 6546-1	32.1
84140 Prostate Cancer (OD04410)	8.3
84141 Prostate NAT (OD04410)	69.3
87073 Prostate Cancer (OD04720-01)	11.7
87074 Prostate NAT (OD04720-02)	40.3
Normal Lung GENPAK 061010	47
83239 Lung Met to Muscle (ODO4286)	0
83240 Muscle NAT (ODO4286)	2.2
84136 Lung Malignant Cancer (OD03126)	31
84137 Lung NAT (OD03126)	21.9
84871 Lung Cancer (OD04404)	3.2
84872 Lung NAT (OD04404)	24.5
84875 Lung Cancer (OD04565)	3.1
84876 Lung NAT (OD04565)	7.9
85950 Lung Cancer (OD04237-01)	37.9
85970 Lung NAT (OD04237-02)	15.6
83255 Ocular Mel Met to Liver (ODO4310)	0
83256 Liver NAT (ODO4310)	12.4
84139 Melanoma Mets to Lung (OD04321)	2.5
84138 Lung NAT (OD04321)	47.3
Normal Kidney GENPAK 061008	13.3
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0
83787 Kidney NAT (OD04338)	11.5
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	1.3
83789 Kidney NAT (OD04339)	5.6
83790 Kidney Ca, Clear cell type (OD04340)	4.2
83791 Kidney NAT (OD04340)	18.3
83792 Kidney Ca, Nuclear grade 3 (OD04348)	1.9
83793 Kidney NAT (OD04348)	7.8
87474 Kidney Cancer (OD04622-01)	4.1
87475 Kidney NAT (OD04622-03)	3.8
85973 Kidney Cancer (OD04450-01)	2.1
85974 Kidney NAT (OD04450-03)	8.3
Kidney Cancer Clontech 8120607	0
Kidney NAT Clontech 8120608	3.7
Kidney Cancer Clontech 8120613	0
Kidney NAT Clontech 8120614	7.2
Kidney Cancer Clontech 9010320	3.5
Kidney NAT Clontech 9010321	5
Normal Uterus GENPAK 061018	0
Uterus Cancer GENPAK 064011	5.9
Normal Thyroid Clontech A+ 6570-1	54.3
Thyroid Cancer GENPAK 064010	7
Thyroid Cancer INVITROGEN A302152	9.9
Thyroid NAT INVITROGEN A302153	32.3
Normal Breast GENPAK 061019	76.3
84877 Breast Cancer (OD04566)	1.3
85975 Breast Cancer (OD04590-01)	2.2
85976 Breast Cancer Mets (OD04590-03)	10.2
87070 Breast Cancer Metastasis (OD04655-05)	2.1

GENPAK Breast Cancer 064006	16.5
Breast Cancer Res. Gen. 1024	69.3
Breast Cancer Clontech 9100266	32.3
Breast NAT Clontech 9100265	45.7
Breast Cancer INVITROGEN A209073	39.5
Breast NAT INVITROGEN A2090734	100
Normal Liver GENPAK 061009	6.2
Liver Cancer GENPAK 064003	3.1
Liver Cancer Research Genetics RNA 1025	4.1
Liver Cancer Research Genetics RNA 1026	5.6
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	8.4
Paired Liver Tissue Research Genetics RNA 6004-N	3.5
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	9.9
Paired Liver Tissue Research Genetics RNA 6005-N	7
Normal Bladder GENPAK 061001	2.4
Bladder Cancer Research Genetics RNA 1023	12.5
Bladder Cancer INVITROGEN A302173	4.3
87071 Bladder Cancer (OD04718-01)	6.7
87072 Bladder Normal Adjacent (OD04718-03)	2.1
Normal Ovary Res. Gen.	7.5
Ovarian Cancer GENPAK 064008	84.7
87492 Ovary Cancer (OD04768-07)	1.4
87493 Ovary NAT (OD04768-08)	2.5
Normal Stomach GENPAK 061017	9.9
Gastric Cancer Clontech 9060358	2.2
NAT Stomach Clontech 9060359	2.3
Gastric Cancer Clontech 9060395	84.7
NAT Stomach Clontech 9060394	25.5
Gastric Cancer Clontech 9060397	17
NAT Stomach Clontech 9060396	2.6
Gastric Cancer GENPAK 064005	3.8

NOV4

Expression of gene NOV4 was assessed using the primer-probe set Ag273, described in Table 20. Results from RTQ-PCR runs are shown in Tables 21 and 22.

Table 20. Probe and Primer Ag273

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CGGCTTGACGATGCTTCAC-3'		19		62
Probe	FAM-5'-TGACTTTTCTGGGCTTACCAATGCTATTTCAA-3'-TAMRA		32		63
Reverse	5'-GCACCTATCTCAATATCTGCAATATT		27		64

	G-3'				

Table 21: Panel 1

Tissue Name	Rel. Expr., % tm379f	Rel. Expr., % tm444f	Rel. Expr., % tm566f_ag273b
Endothelial cells	0	0	0
Endothelial cells (treated)	0	0	0
Pancreas	0	0	0
Pancreatic ca. CAPAN 2	0	0	0
Adipose	0	1.1	26.6
Adrenal gland	0	0	0
Thyroid	0	0	0
Salavary gland	10	14	12.9
Pituitary gland	0	0	0
Brain (fetal)	0	0.2	0
Brain (whole)	0	0.2	0.2
Brain (amygdala)	0	0	0
Brain (cerebellum)	2.6	11.3	1.6
Brain (hippocampus)	0	0	0
Brain (substantia nigra)	0	0.2	0
Brain (thalamus)	0	2.3	2.9
Brain (hypothalamus)	0	0	0
Spinal cord	0	0	0
CNS ca. (glio/astro) U87-MG	0	0	0
CNS ca. (glio/astro) U-118-MG	0	0	0
CNS ca. (astro) SW1783	0	0	0
CNS ca.* (neuro; met) SK-N-AS	2.7	5	6.6
CNS ca. (astro) SF-539	0	0	0
CNS ca. (astro) SNB-75	3.4	16.3	10.2
CNS ca. (glio) SNB-19	21.5	24.1	24.3
CNS ca. (glio) U251	0.2	2.2	4.2
CNS ca. (glio) SF-295	19.9	22.7	37.6
Heart	0	0.8	1.5
Skeletal muscle	0	0	0
Bone marrow	0	0.3	0
Thymus	0	0	0.4
Spleen	0	0	0
Lymph node	0	0	0
Colon (ascending)	1	8.6	9.9
Stomach	3.1	6	0.4
Small Intestine	2.1	5.7	4.2
Colon ca. SW480	0	0	0
Colon ca.* (SW480 met)SW620	0	0	0
Colon ca. HT29	12	10.4	34.4
Colon ca. HCT-116	0	0	0
Colon ca. CaCo-2	0	0	0
Colon ca. HCT-15	0	0	0

Colon ca. HCC-2998	0	0	0
Gastric ca.* (liver met) NCI-N87	3.1	6.9	1.3
Bladder	2.4	14	0.1
Trachea	0.4	1.7	8.9
Kidney	0	0	0.2
Kidney (fetal)	0	1.5	1.3
Renal ca. 786-0	0	0	0
Renal ca. A498	0	0	0
Renal ca. RXF 393	0	0	0
Renal ca. ACHN	0	0	0
Renal ca. UO-31	0	0	0
Renal ca. TK-10	0	0	0
Liver	0.1	2.3	0
Liver (fetal)	0	0.8	0
Liver ca. (hepatoblast) HepG2	0	0	0
Lung	0	2	0.5
Lung (fetal)	0.9	6.7	2.2
Lung ca. (small cell) LX-1	0	0	0
Lung ca. (small cell) NCI-H69	0	1.8	2.7
Lung ca. (s.cell var.) SHP-77	100	100	44.1
Lung ca. (large cell)NCI-H460	0	0	0
Lung ca. (non-sm. cell) A549	0	0.4	0
Lung ca. (non-s.cell) NCI-H23	0	5.2	14.7
Lung ca (non-s.cell) HOP-62	0	2.5	12.2
Lung ca. (non-s.cl) NCI-H522	0	0	0.2
Lung ca. (squam.) SW 900	8.4	9.8	11.9
Lung ca. (squam.) NCI-H596	0	1.9	2.5
Mammary gland	0	1.3	4.8
Breast ca.* (pl. effusion) MCF-7	0	0.2	0.4
Breast ca.* (pl.ef) MDA-MB-231	0	0	0
Breast ca.* (pl. effusion) T47D	0.1	4.6	7.2
Breast ca. BT-549	0	0.7	0
Breast ca. MDA-N	0	0	0
Ovary	0	0	0
Ovarian ca. OVCAR-3	0	0	0
Ovarian ca. OVCAR-4	0	0	0
Ovarian ca. OVCAR-5	8.8	7.2	6.2
Ovarian ca. OVCAR-8	0	0	0
Ovarian ca. IGROV-1	0	0	0
Ovarian ca.* (ascites) SK-OV-3	0	0	0
Uterus	0	0	0
Placenta	0	0.2	0.8
Prostate	2.8	5.2	3.6
Prostate ca.* (bone met)PC-3	24.5	21.9	100
Testis	0	0.4	0
Melanoma Hs688(A).T	0	0	0
Melanoma* (met) Hs688(B).T	0	0	0
Melanoma UACC-62	1.2	2.7	0.3
Melanoma M14	0	0	0
Melanoma LOX IMVI	0	0	0

Melanoma* (met) SK-MEL-5	0	0	0
Melanoma SK-MEL-28	0	0	0.2

Table 22: Panel 2D

Tissue Name	Rel. Expr., % 2Dtm2301f_ag273	Rel. Expr., % 2Dtm3156f_ag273
Normal Colon GENPAK 061003	13.4	14.4
83219 CC Well to Mod Diff (ODO3866)	0.2	0.2
83220 CC NAT (ODO3866)	2.9	1.5
83221 CC Gr.2 rectosigmoid (ODO3868)	0	0
83222 CC NAT (ODO3868)	0.2	0.2
83235 CC Mod Diff (ODO3920)	0.3	0.2
83236 CC NAT (ODO3920)	0.8	0.6
83237 CC Gr.2 ascend colon (ODO3921)	3.3	2.7
83238 CC NAT (ODO3921)	1.8	3
83241 CC from Partial Hepatectomy (ODO4309)	0	0.2
83242 Liver NAT (ODO4309)	0.2	0.4
87472 Colon mets to lung (OD04451-01)	0	0.2
87473 Lung NAT (OD04451-02)	2	1.3
Normal Prostate Clontech A+ 6546-1	7.5	4
84140 Prostate Cancer (OD04410)	2.8	2.2
84141 Prostate NAT (OD04410)	7.7	8.4
87073 Prostate Cancer (OD04720-01)	5.7	6.4
87074 Prostate NAT (OD04720-02)	17.9	18.2
Normal Lung GENPAK 061010	4	4.2
83239 Lung Met to Muscle (ODO4286)	0	0.2
83240 Muscle NAT (ODO4286)	0	0
84136 Lung Malignant Cancer (OD03126)	11	8.9
84137 Lung NAT (OD03126)	2.1	2.3
84871 Lung Cancer (OD04404)	19.9	21.6
84872 Lung NAT (OD04404)	3.5	1.6
84875 Lung Cancer (OD04565)	0.6	1
84876 Lung NAT (OD04565)	0.5	0.6
85950 Lung Cancer (OD04237-01)	21.9	14.4
85970 Lung NAT (OD04237-02)	1.4	1.2
83255 Ocular Mel Met to Liver (ODO4310)	0	0
83256 Liver NAT (ODO4310)	0.4	0.3
84139 Melanoma Mets to Lung (OD04321)	0.6	0.6
84138 Lung NAT (OD04321)	3.3	2.3
Normal Kidney GENPAK 061008	0.2	0.2
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0	0
83787 Kidney NAT (OD04338)	0.3	0.3
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0	0
83789 Kidney NAT (OD04339)	0	0.2
83790 Kidney Ca, Clear cell type (OD04340)	0.6	0.3
83791 Kidney NAT (OD04340)	0.2	0.1
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0	0

83793 Kidney NAT (OD04348)	0	0.2
87474 Kidney Cancer (OD04622-01)	0.4	0.4
87475 Kidney NAT (OD04622-03)	0	0
85973 Kidney Cancer (OD04450-01)	0	0
85974 Kidney NAT (OD04450-03)	0.2	0
Kidney Cancer Clontech 8120607	0.4	0.6
Kidney NAT Clontech 8120608	0	0
Kidney Cancer Clontech 8120613	0	0
Kidney NAT Clontech 8120614	0	0
Kidney Cancer Clontech 9010320	0	0
Kidney NAT Clontech 9010321	0	0
Normal Uterus GENPAK 061018	0.6	0
Uterus Cancer GENPAK 064011	0.8	0.6
Normal Thyroid Clontech A+ 6570-1	0.9	0.3
Thyroid Cancer GENPAK 064010	0.1	0.1
Thyroid Cancer INVITROGEN A302152	0	0
Thyroid NAT INVITROGEN A302153	0.6	0.6
Normal Breast GENPAK 061019	11.4	7
84877 Breast Cancer (OD04566)	0.8	0.6
85975 Breast Cancer (OD04590-01)	5.1	3.9
85976 Breast Cancer Mets (OD04590-03)	2.9	1.6
87070 Breast Cancer Metastasis (OD04655-05)	100	100
GENPAK Breast Cancer 064006	3.9	2.7
Breast Cancer Res. Gen. 1024	1.1	0.5
Breast Cancer Clontech 9100266	6.2	3.5
Breast NAT Clontech 9100265	5.2	4
Breast Cancer INVITROGEN A209073	0.9	0.7
Breast NAT INVITROGEN A2090734	2	1.2
Normal Liver GENPAK 061009	5.9	1.7
Liver Cancer GENPAK 064003	0	0
Liver Cancer Research Genetics RNA 1025	0.2	0.2
Liver Cancer Research Genetics RNA 1026	0	0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.4	0
Paired Liver Tissue Research Genetics RNA 6004-N	0	0
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0	0
Paired Liver Tissue Research Genetics RNA 6005-N	0	0
Normal Bladder GENPAK 061001	0.3	0.2
Bladder Cancer Research Genetics RNA 1023	3.9	2.8
Bladder Cancer INVITROGEN A302173	1.5	1.2
87071 Bladder Cancer (OD04718-01)	0.1	0
87072 Bladder Normal Adjacent (OD04718-03)	6.2	4.6
Normal Ovary Res. Gen.	0	0
Ovarian Cancer GENPAK 064008	1	1.2
87492 Ovary Cancer (OD04768-07)	0	0
87493 Ovary NAT (OD04768-08)	0	0
Normal Stomach GENPAK 061017	1.2	1.7
Gastric Cancer Clontech 9060358	0	0
NAT Stomach Clontech 9060359	0.2	0.4
Gastric Cancer Clontech 9060395	1.1	1.1
NAT Stomach Clontech 9060394	0.4	0.3

Gastric Cancer Clontech 9060397	0.2	0.3
NAT Stomach Clontech 9060396	0	0
Gastric Cancer GENPAK 064005	1	1.9

NOV5

Expression of gene NOV5 was assessed using the primer-probe set Ag819, described in Table 23. Results from RTQ-PCR runs are shown in Tables 12, 13, 14, 15 and 16.

Table 23. Probe and Primer Ag819

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GGTCCAACAGGGCTATCAAT-3'	58.9	20	1105	65
Probe	TET-5'-CCAAACCACGACTGTCGTAGCAGGTA-3'-TAMRA	69.1	26	1156	66
Reverse	5'-GCACCTATCTCAATATCTGCAATATTG-3'	59.5	21	1182	67

Table 24. Panel 1.2

Tissue Name	Rel. Expr., % 1.2tm959t_ag819	Rel. Expr., % 1.2tm1100t_ag819
Endothelial cells	0	0
Heart (fetal)	0.4	0.8
Pancreas	43.8	48.3
Pancreatic ca. CAPAN 2	8.1	17.9
Adrenal Gland (new lot*)	0.2	0.2
Thyroid	11.8	12.9
Salavary gland	63.3	63.7
Pituitary gland	0.9	0.5
Brain (fetal)	37.1	41.5
Brain (whole)	4.6	6.3
Brain (amygdala)	1.5	2.3
Brain (cerebellum)	0.9	1.5
Brain (hippocampus)	3.4	4
Brain (thalamus)	1.9	2.6
Cerebral Cortex	1.2	2.7
Spinal cord	1.2	1.9
CNS ca. (glio/astro) U87-MG	0	0
CNS ca. (glio/astro) U-118-MG	0	0
CNS ca. (astro) SW1783	0	0
CNS ca.* (neuro; met) SK-N-AS	0.3	0
CNS ca. (astro) SF-539	0	0
CNS ca. (astro) SNB-75	0	0
CNS ca. (glio) SNB-19	0	0

CNS ca. (glio) U251	0	0.1
CNS ca. (glio) SF-295	0	0
Heart	8.1	9.5
Skeletal Muscle (new lot*)	2.6	3.7
Bone marrow	0.6	1.2
Thymus	0	0
Spleen	0.5	0
Lymph node	1.4	0.2
Colorectal	0.3	1.8
Stomach	10.7	23.3
Small intestine	10.4	18.9
Colon ca. SW480	0	0
Colon ca.* (SW480 met)SW620	9	11.7
Colon ca. HT29	32.5	40.9
Colon ca. HCT-116	5.9	7.9
Colon ca. CaCo-2	100	100
83219 CC Well to Mod Diff (ODO3866)	4.7	5.4
Colon ca. HCC-2998	2.5	3
Gastric ca.* (liver met) NCI-N87	0	0.2
Bladder	39.2	49.7
Trachea	29.7	34.4
Kidney	27.4	25.7
Kidney (fetal)	17.7	19.1
Renal ca. 786-0	0	0
Renal ca. A498	0	0
Renal ca. RXF 393	0	0
Renal ca. ACHN	0	0
Renal ca. UO-31	1	1.6
Renal ca. TK-10	0	0
Liver	8	3.3
Liver (fetal)	2.8	2.7
Liver ca. (hepatoblast) HepG2	12.8	20.2
Lung	5.7	4.2
Lung (fetal)	9.5	7.4
Lung ca. (small cell) LX-1	39	33.4
Lung ca. (small cell) NCI-H69	7.4	10.5
Lung ca. (s.cell var.) SHP-77	0.5	0.6
Lung ca. (large cell)NCI-H460	0	0
Lung ca. (non-sm. cell) A549	0	0.1
Lung ca. (non-s.cell) NCI-H23	0	0
Lung ca (non-s.cell) HOP-62	0	0.2
Lung ca. (non-s.cl) NCI-H522	0	0
Lung ca. (squam.) SW 900	0.6	0.8
Lung ca. (squam.) NCI-H596	14.6	22.1
Mammary gland	33	46.3
Breast ca.* (pl. effusion) MCF-7	0	0
Breast ca.* (pl.ef) MDA-MB-231	0	0
Breast ca.* (pl. effusion) T47D	0.8	1.3
Breast ca. BT-549	0	0
Breast ca. MDA-N	0.4	0.6

Ovary	4.6	0.2
Ovarian ca. OVCAR-3	3.3	4
Ovarian ca. OVCAR-4	27.9	54
Ovarian ca. OVCAR-5	37.4	51
Ovarian ca. OVCAR-8	0	0
Ovarian ca. IGROV-1	3.2	5.5
Ovarian ca.* (ascites) SK-OV-3	0	0
Uterus	1.4	1.2
Placenta	23.2	22.5
Prostate	2.6	2.7
Prostate ca.* (bone met)PC-3	0	0
Testis	19.8	21.9
Melanoma Hs688(A).T	1.7	0
Melanoma* (met) Hs688(B).T	0.7	0
Melanoma UACC-62	1.8	1.7
Melanoma M14	0	0.2
Melanoma LOX IMVI	0	0
Melanoma* (met) SK-MEL-5	0.5	1
Adipose	0.1	0.2

Table 25. Panel 2D

Tissue Name	Rel. Expr., % 2Dtm2318t_ag819	Rel. Expr., % 2Dtm2649t_ag819
Normal Colon GENPAK 061003	17	20.7
83219 CC Well to Mod Diff (ODO3866)	0.9	5.3
83220 CC NAT (ODO3866)	9.5	6
83221 CC Gr.2 rectosigmoid (ODO3868)	5.8	3.9
83222 CC NAT (ODO3868)	0	0.2
83235 CC Mod Diff (ODO3920)	0.7	0.9
83236 CC NAT (ODO3920)	2.8	2.1
83237 CC Gr.2 ascend colon (ODO3921)	26.2	37.4
83238 CC NAT (ODO3921)	4.4	7
83241 CC from Partial Hepatectomy (ODO4309)	13.1	20.4
83242 Liver NAT (ODO4309)	0.1	0.2
87472 Colon mets to lung (OD04451-01)	8.5	6.2
87473 Lung NAT (OD04451-02)	2.1	1.7
Normal Prostate Clontech A+ 6546-1	1.2	0.3
84140 Prostate Cancer (OD04410)	0.5	0.7
84141 Prostate NAT (OD04410)	0.8	0.6
87073 Prostate Cancer (OD04720-01)	0.4	0.4
87074 Prostate NAT (OD04720-02)	2	2
Normal Lung GENPAK 061010	4.6	4.7
83239 Lung Met to Muscle (ODO4286)	0	0
83240 Muscle NAT (ODO4286)	0.2	0.3
84136 Lung Malignant Cancer (OD03126)	8.7	6.5
84137 Lung NAT (OD03126)	1.4	1.6

84871 Lung Cancer (OD04404)	0	0.2
84872 Lung NAT (OD04404)	3.1	1.2
84875 Lung Cancer (OD04565)	0.2	0
84876 Lung NAT (OD04565)	1	0.9
85950 Lung Cancer (OD04237-01)	100	74.7
85970 Lung NAT (OD04237-02)	1.7	1.5
83255 Ocular Mel Met to Liver (ODO4310)	0.1	0.3
83256 Liver NAT (ODO4310)	0.6	0.2
84139 Melanoma Mets to Lung (OD04321)	14.8	12.2
84138 Lung NAT (OD04321)	1.7	1.5
Normal Kidney GENPAK 061008	23.2	17.4
83786 Kidney Ca, Nuclear grade 2 (OD04338)	4.2	5.1
83787 Kidney NAT (OD04338)	8	11.3
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	65.5	69.3
83789 Kidney NAT (OD04339)	6.7	6
83790 Kidney Ca, Clear cell type (OD04340)	0	0.1
83791 Kidney NAT (OD04340)	13.8	12
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0	0
83793 Kidney NAT (OD04348)	9.2	6.2
87474 Kidney Cancer (OD04622-01)	0.7	0.4
87475 Kidney NAT (OD04622-03)	1.1	1.2
85973 Kidney Cancer (OD04450-01)	32.5	24.5
85974 Kidney NAT (OD04450-03)	22.1	16
Kidney Cancer Clontech 8120607	4.4	4.2
Kidney NAT Clontech 8120608	3.9	2
Kidney Cancer Clontech 8120613	0	0
Kidney NAT Clontech 8120614	1.2	0.8
Kidney Cancer Clontech 9010320	7.7	8
Kidney NAT Clontech 9010321	7.8	6
Normal Uterus GENPAK 061018	0	0
Uterus Cancer GENPAK 064011	24.1	18.8
Normal Thyroid Clontech A+ 6570-1	4.7	2.4
Thyroid Cancer GENPAK 064010	4	2.2
Thyroid Cancer INVITROGEN A302152	0	0
Thyroid NAT INVITROGEN A302153	2.9	2.7
Normal Breast GENPAK 061019	16.6	7.4
84877 Breast Cancer (OD04566)	0.6	0.4
85975 Breast Cancer (OD04590-01)	0.8	0.5
85976 Breast Cancer Mets (OD04590-03)	0	0
87070 Breast Cancer Metastasis (OD04655-05)	0.1	0
GENPAK Breast Cancer 064006	15.7	11.7
Breast Cancer Res. Gen. 1024	12.1	11.6
Breast Cancer Clontech 9100266	1.2	0.6
Breast NAT Clontech 9100265	3	2
Breast Cancer INVITROGEN A209073	6.5	4.6
Breast NAT INVITROGEN A2090734	25	9
Normal Liver GENPAK 061009	0.6	0.5
Liver Cancer GENPAK 064003	0	0
Liver Cancer Research Genetics RNA 1025	0.2	0.3
Liver Cancer Research Genetics RNA 1026	0.2	0.1

Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.2	0.1
Paired Liver Tissue Research Genetics RNA 6004-N	1.6	1.7
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.1	0.2
Paired Liver Tissue Research Genetics RNA 6005-N	0	0.2
Normal Bladder GENPAK 061001	14.8	18.7
Bladder Cancer Research Genetics RNA 1023	6.9	6.4
Bladder Cancer INVITROGEN A302173	0.2	0.1
87071 Bladder Cancer (OD04718-01)	0.1	0
87072 Bladder Normal Adjacent (OD04718-03)	0.2	0.4
Normal Ovary Res. Gen.	0	0
Ovarian Cancer GENPAK 064008	68.8	100
87492 Ovary Cancer (OD04768-07)	0.5	1
87493 Ovary NAT (OD04768-08)	0	0.1
Normal Stomach GENPAK 061017	5	4.5
Gastric Cancer Clontech 9060358	2.5	2.6
NAT Stomach Clontech 9060359	5.6	7
Gastric Cancer Clontech 9060395	1.7	1.3
NAT Stomach Clontech 9060394	3.4	6.4
Gastric Cancer Clontech 9060397	26.1	39.8
NAT Stomach Clontech 9060396	2.7	2.7
Gastric Cancer GENPAK 064005	15.5	22.1

NOV6

Expression of gene NOV6 was assessed using the primer-probe set Ag1395, described in Table 26. Results from RTQ-PCR runs are shown in Tables 12, 13, 14, 15 and 16.

Table 26. Primer and Probe Ag1395

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CCTCCTGCAGGATAAAGTCAT-3'	58.3	21	1518	68
Probe	TET-5'-CCCCAAGGCTCCAGCTACTCTAAATT-3'-TAMRA	66.6	26	1539	69
Reverse	5'-CTCCTGGAGCAGCAATAACTTA-3'	58.7	22	1577	70

Table 27. Panel 1.2

Tissue Name	Rel. Expr., % 1.2tm1618t_ag1389	Rel. Expr., % 1.2tm1729t_ag1389*
Endothelial cells	0	0
Heart (fetal)	3.4	6.8
Pancreas	2.4	0.5
Pancreatic ca. CAPAN 2	0	0

Adrenal Gland (new lot*)	33.9	15.6
Thyroid	2	0.5
Salavary gland	8.7	3.3
Pituitary gland	3.9	1.7
Brain (fetal)	0.2	0.1
Brain (whole)	12.6	0.4
Brain (amygdala)	0.5	0.5
Brain (cerebellum)	4	1.6
Brain (hippocampus)	0.4	0.9
Brain (thalamus)	0.2	0.4
Cerebral Cortex	2.3	3.3
Spinal cord	0.2	0.1
CNS ca. (glio/astro) U87-MG	0	0
CNS ca. (glio/astro) U-118-MG	0	0
CNS ca. (astro) SW1783	1.2	1
CNS ca.* (neuro; met) SK-N-AS	0.5	0.1
CNS ca. (astro) SF-539	3.8	3.1
CNS ca. (astro) SNB-75	0	0
CNS ca. (glio) SNB-19	0	0
CNS ca. (glio) U251	0	0
CNS ca. (glio) SF-295	0	0.3
Heart	6.7	29.5
Skeletal Muscle (new lot*)	3.6	9.2
Bone marrow	0.3	0.5
Thymus	0.7	0.2
Spleen	3.4	2.3
Lymph node	0.6	0.2
Colorectal	0.3	0.5
Stomach	4.1	1.8
Small Intestine	18.9	17.1
Colon ca. SW480	0	0.3
Colon ca.* (SW480 met)SW620	1.8	1.9
Colon ca. HT29	0	0
Colon ca. HCT-116	0.2	0.2
Colon ca. CaCo-2	0.2	0.2
83219 CC Well to Mod Diff (ODO3866)	5.2	2.3
Colon ca. HCC-2998	1	1.5
Gastric ca.* (liver met) NCI-N87	13.8	6.2
Bladder	12.2	15.5
Trachea	0.8	0.5
Kidney	6.1	9.6
Kidney (fetal)	0.5	1.8
Renal ca. 786-0	0	0
Renal ca. A498	0.1	0.2
Renal ca. RXF 393	4.5	6.8
Renal ca. ACHN	0	0.2
Renal ca. UO-31	2.4	7
Renal ca. TK-10	1.2	2.1
Liver	3.5	10.9
Liver (fetal)	2.9	5.3

Liver ca. (hepatoblast) HepG2	2.6	1.8
Lung	0.7	0.4
Lung (fetal)	0.9	2.8
Lung ca. (small cell) LX-1	4.8	6.5
Lung ca. (small cell) NCI-H69	0.1	0.2
Lung ca. (s.cell var.) SHP-77	0	0
Lung ca. (large cell)NCI-H460	0.7	1.6
Lung ca. (non-sm. cell) A549	0.2	0.4
Lung ca. (non-s.cell) NCI-H23	1.3	3.4
Lung ca (non-s.cell) HOP-62	1.9	10.6
Lung ca. (non-s.cl) NCI-H522	1.4	3.2
Lung ca. (squam.) SW 900	0.5	0.8
Lung ca. (squam.) NCI-H596	0	0
Mammary gland	76.8	7.3
Breast ca.* (pl. effusion) MCF-7	0.5	0.2
Breast ca.* (pl.ef) MDA-MB-231	0.5	0.4
Breast ca.* (pl. effusion) T47D	0.3	0.2
Breast ca. BT-549	100	55.9
Breast ca. MDA-N	0.2	0.3
Ovary	11.1	19.9
Ovarian ca. OVCAR-3	0.1	0.3
Ovarian ca. OVCAR-4	0	0
Ovarian ca. OVCAR-5	0.6	0.7
Ovarian ca. OVCAR-8	4.1	1.7
Ovarian ca. IGROV-1	0.1	0
Ovarian ca.* (ascites) SK-OV-3	1.2	1.4
Uterus	13	19.8
Placenta	3.9	1.3
Prostate	67.4	100
Prostate ca.* (bone met)PC-3	0	0
Testis	0.5	0.2
Melanoma Hs688(A).T	2.9	8.8
Melanoma* (met) Hs688(B).T	1.1	3.1
Melanoma UACC-62	0.2	0.3
Melanoma M14	10.4	42.6
Melanoma LOX IMVI	0.1	0.4
Melanoma* (met) SK-MEL-5	0	0.1
Adipose	3.6	6.6

Table 28. Panel 2D

Tissue Name	Rel. Expr., % 2Dtm2491t_ag1389	Rel. Expr., % 2Dtm2507t_ag1389
Normal Colon GENPAK 061003	1	1.8
83219 CC Well to Mod Diff (ODO3866)	1.6	3.1
83220 CC NAT (ODO3866)	0.5	0.5
83221 CC Gr.2 rectosigmoid (ODO3868)	0.4	0.6

83222 CC NAT (ODO3868)	0.2	0.2
83235 CC Mod Diff (ODO3920)	0.3	0.3
83236 CC NAT (ODO3920)	0.4	0.5
83237 CC Gr.2 ascend colon (ODO3921)	1.7	1.4
83238 CC NAT (ODO3921)	0.7	0.5
83241 CC from Partial Hepatectomy (ODO4309)	1.1	0.5
83242 Liver NAT (ODO4309)	0.8	0.5
87472 Colon mets to lung (OD04451-01)	0.2	0.2
87473 Lung NAT (OD04451-02)	0.2	0.4
Normal Prostate Clontech A+ 6546-1	4.9	36.6
84140 Prostate Cancer (OD04410)	100	100
84141 Prostate NAT (OD04410)	14.9	10.9
87073 Prostate Cancer (OD04720-01)	1.1	1
87074 Prostate NAT (OD04720-02)	2.9	2
Normal Lung GENPAK 061010	0.6	0.7
83239 Lung Met to Muscle (ODO4286)	0.3	0.1
83240 Muscle NAT (ODO4286)	0.3	0.5
84136 Lung Malignant Cancer (OD03126)	1.9	1.4
84137 Lung NAT (OD03126)	0.7	0.7
84871 Lung Cancer (OD04404)	0.8	0.5
84872 Lung NAT (OD04404)	0.7	1.5
84875 Lung Cancer (OD04565)	0.5	0.4
84876 Lung NAT (OD04565)	0.1	0.3
85950 Lung Cancer (OD04237-01)	2.2	2.5
85970 Lung NAT (OD04237-02)	0.9	1.4
83255 Ocular Mel Met to Liver (ODO4310)	0	0
83256 Liver NAT (ODO4310)	1.1	0.6
84139 Melanoma Mets to Lung (OD04321)	0.8	0.3
84138 Lung NAT (OD04321)	1.5	0.4
Normal Kidney GENPAK 061008	0.8	0.4
83786 Kidney Ca, Nuclear grade 2 (OD04338)	7	3.7
83787 Kidney NAT (OD04338)	0.8	0.4
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	2.6	5.9
83789 Kidney NAT (OD04339)	0.4	0.5
83790 Kidney Ca, Clear cell type (OD04340)	0.7	0.5
83791 Kidney NAT (OD04340)	0.6	0.9
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.4	0.3
83793 Kidney NAT (OD04348)	0.6	0.5
87474 Kidney Cancer (OD04622-01)	8.5	8.4
87475 Kidney NAT (OD04622-03)	0.2	0.2
85973 Kidney Cancer (OD04450-01)	10.7	5.1
85974 Kidney NAT (OD04450-03)	0.5	0.3
Kidney Cancer Clontech 8120607	0.2	0.2
Kidney NAT Clontech 8120608	0.4	0.3
Kidney Cancer Clontech 8120613	0.4	0.3
Kidney NAT Clontech 8120614	0.4	0.2
Kidney Cancer Clontech 9010320	1.7	5
Kidney NAT Clontech 9010321	2.1	3.3
Normal Uterus GENPAK 061018	2	1.6
Uterus Cancer GENPAK 064011	1.6	1.2

Normal Thyroid Clontech A+ 6570-1	0.3	1.8
Thyroid Cancer GENPAK 064010	0	0
Thyroid Cancer INVITROGEN A302152	0.1	0
Thyroid NAT INVITROGEN A302153	0.5	0.2
Normal Breast GENPAK 061019	1.8	1.4
84877 Breast Cancer (OD04566)	0.2	0.4
85975 Breast Cancer (OD04590-01)	1	2.2
85976 Breast Cancer Mets (OD04590-03)	1.1	2.9
87070 Breast Cancer Metastasis (OD04655-05)	0.2	0.2
GENPAK Breast Cancer 064006	1	0.9
Breast Cancer Res. Gen. 1024	2	5.8
Breast Cancer Clontech 9100266	0.4	0.4
Breast NAT Clontech 9100265	0.9	1.1
Breast Cancer INVITROGEN A209073	1.3	1.4
Breast NAT INVITROGEN A2090734	0.9	0.5
Normal Liver GENPAK 061009	0.2	0.3
Liver Cancer GENPAK 064003	1.1	2.5
Liver Cancer Research Genetics RNA 1025	0.2	0.3
Liver Cancer Research Genetics RNA 1026	4.1	3.2
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.4	1.5
Paired Liver Tissue Research Genetics RNA 6004-N	0.6	1.7
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	3.8	7.2
Paired Liver Tissue Research Genetics RNA 6005-N	0.7	1.2
Normal Bladder GENPAK 061001	1.6	2.4
Bladder Cancer Research Genetics RNA 1023	0.2	0.2
Bladder Cancer INVITROGEN A302173	0.2	0.1
87071 Bladder Cancer (OD04718-01)	1	1.2
87072 Bladder Normal Adjacent (OD04718-03)	0.5	1.4
Normal Ovary Res. Gen.	1	1.7
Ovarian Cancer GENPAK 064008	3.5	3
87492 Ovary Cancer (OD04768-07)	0.1	0.4
87493 Ovary NAT (OD04768-08)	1	1.2
Normal Stomach GENPAK 061017	0.3	0.8
Gastric Cancer Clontech 9060358	0.2	0.5
NAT Stomach Clontech 9060359	0.3	0.9
Gastric Cancer Clontech 9060395	1.1	2.7
NAT Stomach Clontech 9060394	1	1.6
Gastric Cancer Clontech 9060397	2.8	10.6
NAT Stomach Clontech 9060396	0.2	0.6
Gastric Cancer GENPAK 064005	0.9	2

Example 2: SAGE analysis for NOVX

Serial Analysis of Gene Expression, or SAGE, is an experimental technique designed to gain a quantitative measure of gene expression. The SAGE technique itself includes several steps utilizing molecular biological, DNA sequencing and bioinformatics techniques. These steps (reviewed in Adams MD, "Serial analysis of gene expression: ESTs get smaller." Bioessays. 18(4):261-2 (1996)) have been used to produce 9 or 10 base "tags", which are then, in some manner, assigned gene descriptions. For experimental reasons, these tags are immediately adjacent to the 3' end of the 3'-most NlaIII restriction site in cDNA sequences. The Cancer Genome Anatomy Project, or CGAP, is an NCI-initiated and sponsored project, which hopes to delineate the molecular fingerprint of the cancer cell. It has created a database of those cancer-related projects that used SAGE analysis in order to gain insight into the initiation and development of cancer in the human body. The SAGE expression profiles reported in this invention are generated by first identifying the Unigene accession ID associated with the given MTC gene by querying the Unigene database at http://www.ncbi.nlm.nih.gov/UniGene/. This page has then a link to the SAGE : Gene to Tag mapping (http://www.ncbi.nlm.nih.gov/SAGE/SAGEcid.cgi?cid="unigeneID").

This generated the reports that are included in this application, which list the number of tags found for the given gene in a given sample along with the relative expression. This information is then used to understand whether the gene has a more general role in tumorogenesis and/or tumor progression. A list of the SAGE libraries generated by CGAP and used in the analysis can be found at http://www.ncbi.nlm.nih.gov/SAGE/sagelb.cgi.

NOV2 SAGE Data

CGAGAGGGAG	Library name	Tags per million	Tag counts	Total tags
	SAGE Duke BB542 normal cerebellum	313	13	41450
	SAGE Duke cerebellum	434	3	6899
	SAGE Panc 91-16113	29	1	34159
	SAGE OVT-7	18	1	55476
	SAGE normal cerebellum	399	18	45079
	SAGE Duke H1043	25	2	77449
GCATTTGCAG	Library name	Tags per million	Tag counts	Total tags
	SAGE normal pool(6th)	15	1	64136
	SAGE normal cerebellum	22	1	45079
	SAGE ML10-10	17	1	57326
	SAGE IOSE29-11	20	1	48876

NOV3 SAGE Data

SAGE library data and reliable tag summary:

Reliable tags found in SAGE libraries:

CATAAAGACT	Tags per	Tag	
Library name	*million*	*counts*	*Total tags*
SAGE Duke thalamus	40	1	24671
SAGE Chen LNCaP	15	1	62681
SAGE Duke GBM H1110	14	1	71138
SAGE SW837	16	1	61290
SAGE Tu102	17	1	58190
SAGE OVT-8	23	1	43074
SAGE H1126	56	1	17756
SAGE normal cerebellum	22	1	45079
SAGE OVT-8	117	4	34096

NOV4 SAGE Data

UniGene cluster: Hs. 255372 Submit

Hs.255372 : **hypothetical protein DKFZp564O1278**

SAGE library data and reliable tag summary:

Reliable tags found in SAGE libraries:

CTGAACCTGA	Tags per million	Tag counts	Total tags
Library name			
SAGE HCT116	16	1	60322
SAGE Caco 2	16	1	61601
SAGE Chen Tumor Pr	14	1	68384
SAGE HX	93	3	32157
SAGE H126	185	6	32420
SAGE Duke H392	17	1	57529
SAGE SW837	16	1	60986
SAGE RKO	96	5	52064
SAGE PR317 normal prostate	16	1	59419
SAGE NC1	19	1	50115
SAGE Tu98	61	3	49005
SAGE SciencePark MCF7 Control 0h	16	1	61079
SAGE LNCaP	44	1	22637
SAGE OVT-7	18	1	54914
SAGE MDA453	52	1	18924
SAGE mammary epithelium	20	1	49167
SAGE OVT-8	29	1	33575
SAGE Duke-H988	35	1	28015

Reliable tags **NOT** found in SAGE libraries:

NOV5 SAGE Data

TGCAGATTGC Library name	Tags per million	Tag counts	Total tags
SAGE Caco 2	32	2	61888
SAGE Duke GBM H1110	14	1	71138
SAGE SW837	16	1	61280
SAGE NHA(5th)	18	1	53218
SAGE NC2	88	5	50128
SAGE Panc 81-16113	58	2	34158
SAGE Panc 98-8252	27	1	36067
SAGE Tu102	34	2	58180
SAGE Tu98	40	2	48527
SAGE Duke H341	66	3	44883
SAGE OVT-8	23	1	43074
SAGE OVT-7	18	1	55476
SAGE mammary epithelium	40	2	49782
SAGE DCIS	144	6	41540
SAGE OVT-8	58	2	34086
SAGE Duke 98-349	705	4	5669
SAGE Duke-H988	35	1	28103
SAGE DCIS 2	34	1	29201
SAGE Br N	28	1	36274
SAGE Duke H1043	12	1	77449

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
   (a) a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12;
   (b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;
   (c) an amino acid sequence selected from the group consisting SEQ ID NOS:2, 4, 6, 8, 10, and 12; and
   (d) a variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence.

2 The polypeptide of claim 1, wherein said polypeptide comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence selected from the group consisting SEQ ID NOS:2, 4, 6, 8, 10, and 12.

3. The polypeptide of claim 2, wherein said allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, and 11.

4. The polypeptide of claim 1, wherein the amino acid sequence of said variant comprises a conservative amino acid substitution.

5. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12;

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12;

(d) a variant of an amino acid sequence selected from the group consisting SEQ ID NOS:2, 4, 6, 8, 10, and 12, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence;

(e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence chosen from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, and 12, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; and

(f) a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, and 11.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, and 11;

(b) a nucleotide sequence differing by one or more nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, and 11, provided that no more than 20% of the nucleotides differ from said nucleotide sequence;

(c) a nucleic acid fragment of (a); and

(d) a nucleic acid fragment of (b).

10. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to a nucleotide sequence chosen from the group consisting SEQ ID NOS:1, 3, 5, 7, 9, and 11, or a complement of said nucleotide sequence.

11. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence;

(b) an isolated second polynucleotide that is a complement of the first polynucleotide; and

(c) a nucleic acid fragment of (a) or (b).

12. A vector comprising the nucleic acid molecule of claim 11.

13. The vector of claim 12, further comprising a promoter operably-linked to said nucleic acid molecule.

14. A cell comprising the vector of claim 12.

15. An antibody that binds immunospecifically to the polypeptide of claim 1.

16. The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17. The antibody of claim 15, wherein the antibody is a humanized antibody.

18. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of the probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

20. The method of claim 19 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

21. The method of claim 20 wherein the cell or tissue type is cancerous.

22. A method of identifying an agent that binds to a polypeptide of claim 1, the method comprising:

(a) contacting said polypeptide with said agent; and

(b) determining whether said agent binds to said polypeptide.

23. The method of claim 22 wherein the agent is a cellular receptor or a downstream effector.

24. A method for identifying an agent that modulates the expression or activity of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing said polypeptide;

(b) contacting the cell with said agent, and

(c) determining whether the agent modulates expression or activity of said polypeptide,

whereby an alteration in expression or activity of said peptide indicates said agent modulates expression or activity of said polypeptide.

25. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of said claim with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

26. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the polypeptide of claim 1 in an amount sufficient to treat or prevent said NOVX-associated disorder in said subject.

27. The method of claim 26 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

28. The method of claim 26 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

29. The method of claim 26, wherein said subject is a human.

30. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the nucleic acid of claim 5 in an amount sufficient to treat or prevent said NOVX-associated disorder in said subject.

31. The method of claim 30 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

32. The method of claim 30 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

33. The method of claim 30, wherein said subject is a human.

34. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the antibody of claim 15 in an amount sufficient to treat or prevent said NOVX-associated disorder in said subject.

35. The method of claim 34 wherein the disorder is diabetes.

36. The method of claim 34 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

37. The method of claim 34, wherein the subject is a human.

38. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically-acceptable carrier.

39. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically-acceptable carrier.

40. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically-acceptable carrier.

41. A kit comprising in one or more containers, the pharmaceutical composition of claim 38.

42. A kit comprising in one or more containers, the pharmaceutical composition of claim 39.

43. A kit comprising in one or more containers, the pharmaceutical composition of claim 40.

44. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

(a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

(b) comparing the amount of said polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease;

wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

45. The method of claim 44 wherein the predisposition is to a cancer.

46. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 5 in a first mammalian subject, the method comprising:

(a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and

(b) comparing the amount of said nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

47. The method of claim 46 wherein the predisposition is to a cancer.

48. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising an amino acid sequence of at least one of SEQ ID NOS:2, 4, 6, 8, 10, and 12, or a biologically active fragment thereof.

49. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the pathological state.

50. The method of claim 49, wherein the pathological state is cancer related.